Structural basis of light dependent modulation of phytochrome

Soshichiro Nagano

School of Biological and Chemical Sciences

Queen Mary, University of London

A thesis submitted for the degree of Doctor of Philosophy

Statement of originality

This thesis includes the ideas originally developed by a collaboration consisting mainly of Dr Norbert Krauß, Professor Tilman Lamparter, and Professor Katsuhiko Inomata. No conflict of interest is declared with regards to the works shown in this thesis, and I certify that any additional ideas and all the experimental data shown herein are my original works, unless otherwise credited.

Any other ideas or data from other people, published or otherwise are fully acknowledged according to the standard referencing practices.

Soshichiro Nagano

Abstract

This thesis presents structural and biochemical studies of a phytochrome, Agp1, a bilin-binding red-light receptor protein. Crystallographic studies were undertaken in order to obtain structural insight into the mechanism of photoconversion and signal transduction from the sensor domain to the signalling domain, the latter hypothesised to be universally conserved among two-component histidine kinases. Using Agp1 from a non-photosynthetic plant pathogen, *Agrobacterium tumefaciens* as a model phytochrome, structural determination of both Pr and Pfr-forms was attempted, in order to permit unbiased structural comparison of the two. Application of the surface entropy reduction strategy led to determination of the structure of Agp1 in its Pr-form and crystals of Agp1 in the Pfr-like form were obtained for the first time.

Biochemical studies were undertaken to probe the conformational differences between Agp1 as apoprotein, Pr-form, and Pfr-like form. Limited differences in secondary structure exist between the forms of Agp1. Conformational differences between the Pr and the Pfr-like form seem to underlie the fact that the space group of Agp1 crystals in the Pfr-like form is different from that of Agp1 in the Pr-form. The ability of the Agp1 apoprotein to form a dimer via a disulphide bond at the N-terminal chromophore-binding cysteine residue implies flexibility of the N-terminal region which allows for the initial bilin incorporation during holoprotein formation.

Acknowledgements

First and foremost, I would like to express my utmost gratitude to Dr Norbert Krauß for his supervision. Without his guidance and encouragement, my Ph.D. project would not have been realised and reached the current stage. I would also like to thank his wife Dr Kristina Zubow for the efforts she has paid in setting up the lab as well as invaluable technical advices.

I would also like to thank the Ph.D. panel members, Professor John Allen, Dr James Sullivan, and Dr Robin Maytum for their advices. I thank Professor Richard Pickersgill for the helps I received throughout the project. I thank Dr Jon Nield for offering to include his research results in my thesis.

I am thankful to our collaborators, without whom the project would not have been realised. I thank Dr Patrick Scheerer for technical advices and invaluable suggestions. I thank Professor Tilman Lamparter, Dr Inga Oberpichler, Dr Benjamin Zienicke, Dr Ibrahim Njimona, and Dr Gregor Rottwinkel for their hospitality during my visit to the Karlsruhe Institute of Technology and the plasmids I have received. I would like to thank Professor Katsuhiko Inomata for supplying the precious locked chromophore. I thank Dr David von Stetten for his kind help we received at the ESRF synchrotron, Grenoble.

I am indebted to my friends for their critical opinions, collaborations, and friendships. Thank you, Allan Pang, for being a fantastic lab mate to work with for a long time. I thank Iskander Ibrahim for our collaboration on the chloroplast sensor kinase project. I thank Dennis Nürnberg for our collaboration.

I acknowledge Queen Mary University of London for the collage studentship. The work presented in the present study has been supported by the University of London central research fund. I thank the British Crystallographic Association and International Centre for Diffraction Data for supporting my attendances at the BCA spring meetings. I thank the Collaborative Computational Project No. 4 for awarding the bursary for the CCP4 study weekend 2012.

Finally I would like to thank my parents who taught me the value of perseverance, without which I would not have been able to come this far.

Abbreviations

Abbreviation	Description
a	*anti* conformation of a single bond in a bilin.
Agp1 or 2	*Agrobacterium tumefaciens* phytochrome 1 or 2.
APS	Ammonium persulphate.
AU	Spectrophotometric absorption unit.
BphP	Bacteriophytochromes.
BSA	Bovine serum albumin.
BV	Biliverdin IXα.
CA	Catalytic and ATP-binding domain.
CBD	Chromophore binding domain, a part of the N-terminal signalling domain of phytochromes, consisting of PAS and GAF domains.
Cph1 or 2	*Synechocystis* PCC6803 phytochrome 1 or 2.
DEPC	Diethylpyrocarbonate.
DHp	Dimerisation and histidine phosphorylation domain.
DLS	Dynamic light scattering.
DrBphP	*Deinococcus radiodurans* phytochrome.
DTT	Dithiothreitol.
E	*trans* configuration of a double bond in a bilin.
EPR	Electron paramagnetic resonance.
GAF	cGMP phosphodiesterase/adenylcyclase/FhlA domain.
HEWL	Hen egg white lysozyme.
HIC	Hydrophobic interaction chromatography.
HK	Histidine kinase.
Hpt	Histidine-containing phosphotransfer domain.
HSA	Human serum albumin.
IEX	Ion exchange chromatography.
λ_{max}	The wavelength with the maximum absorption in the absorption spectrum.
M15	Agp1 fragment containing the entire PCM, containing PAS, GAF, and PHY domains (1-504).
M15s	M15 construct without the first nine amino acid residues (10-504).
M20	Agp1 fragment containing the entire CBD, containing PAS and GAF domains (1-313).
M20s	M20 construct without the first nine amino acid residues (10-313).
mAU	Milli absorption unit (spectrophotometric).
NMR	Nuclear magnetic resonance.
PΦB	Phytochromobilin.
PaBphP	*Pseudomonas aeruginosa* phytochrome.
PAS	Per/Arndt/Sim domain.
PCB	Phycocyanobilin.
PCM	Photosensory core module, the entire N-terminal signalling domain of phytochromes, usually consisting of PAS, GAF, and PHY domains.
PEB	Phycoerythrobilin.
Pfr-form	The far-red light absorbing form of phytochromes.

PHY	Phytochrome specific domain.
pI	Isoelectric point.
Pr-form	The red-light absorbing form of phytochromes.
RpBphP1, 2, 3, 4, or 5	*Rhodopseudomonas palustris* phytochrome 1, 2, 3, 4, or 5.
RR	Response regulator.
s	*syn* conformation of a single bond in a bilin.
SAXS	Small angle X-ray scattering.
SDS	Sodium dodecyl sulphate.
SEC	Size exclusion chromatography.
SER	Surface entropy reduction strategy, or prefix of the codename given to constructs modified according to the strategy.
SK	Sensor kinase.
SyB-Cph1	*Synechococcus* OS-B′ phytochrome.
TCS	Two-component signal transduction system.
TEMED	N,N,N′,N′-Tetramethylethylenediamine.
Tm	Melting temperature of a primer (°C).
Tris	Tris(hydroxymethyl)aminomethane.
WT	Wild type.
Z	*cis* configuration of a double bond in a bilin.

Table of contents

List of figures

List of tables

Chapter 1: Introduction

1.1 Photoreception

In the vast majority of ecosystems, light is the sole ultimate energy source, being converted into chemical potential energy by the primary producers. This energy is then distributed to other organisms at higher trophic levels, except in the case of chemoautotrophic organisms. During the course of evolution, life has developed a multitude of implementations to utilise light, for example as an energy source or for sensory purposes. The emergence of photosynthesis in evolution allowed more complex organisms to emerge, and light perception plays vital roles in spatial recognition in both aquatic and terrestrial organisms of a wide range of complexity. Light has been and will continue to be the essential factor for the maintenance of life on the Earth.

Light-sensitive proteins include photosynthetic machineries, opsins, photolyases, cryptochromes, phytochromes and phototropins, to name a few. Obviously research on phytochromes is just one out of many areas of photobiology, but studying them is of intrinsic value and may lead to identification of commonalities with other signalling proteins. Detailed understanding of such common mechanisms will have wide implications.

1.2 Two-component signal transduction systems

1.2.1 Introduction

Two-component signal transduction systems (TCS) provide a ubiquitous mechanism to respond to the environmental stimuli and are found in all domains of life except for animals (Casino et al., 2009). Cellular responses including chemotaxis, osmoregulation, anaerobic respiration, photosynthesis, nitrogen and phosphate acquisition, sporulation, host recognition by pathogens, antibiotic production, the cell cycle, and viability are controlled by histidine kinases (Karniol and Vierstra, 2004, Stock et al., 2000). The name of the system was introduced in 1986 (Nixon et al., 1986) and refers to a conserved group of proteins of two types which together are involved in a phosphotransfer cascade, the activity of which depends on the response to the environmental trigger and modulates the physiological activity (Bourret and Silversmith, 2010). In its most basic form, a TCS consists of a pair of proteins: the sensor kinase (SK) and the response regulator (RR), but more extensive forms of TCS include other

intermediate phosphotransfer proteins (Fig. 1.1). A bacterial genome can encode many TCS to facilitate response to various aspects of the ever-changing environment. A single bacterial cell can contain between 200 and 300 pairs of SK-RR (Laub and Goulian, 2007). The genes of the SK and the cognate RR may be found within the same operon (Szurmant and Hoch, 2010). The interaction between the SK and its cognate response regulator is usually specific (Capra et al., 2010), but cases of promiscuous interaction (Ninfa et al., 1988) or a case where two SKs appears to have redundant roles (Bick et al., 2009) have been reported. The gene of a TCS protein may not be in an operon with the cognate partner, and is then referred to as an "orphan" (Szurmant and Hoch, 2010).

Components in a TCS ultimately translate external stimuli into physiological modulation via a phosphotransfer cascade. In brief, the external stimulus is sensed by the sensor domain of the SK, and the kinase activity of the SK is modulated according to the concentration of the stimulus. Some elaborate TCSs, however, for example the chemotactic signal transduction pathways in *Escherichia coli* and *Salmonella typhimurium*, enable the cell to respond to changes in stimuli concentration. Chemotaxis enables the cell to move towards higher concentration of attractants, but eventual methylation of the membrane receptor protein attenuates the attractant signal to prestimulus levels (p89-97, Hoch and Silhavy, 1995).

Sensor kinases autophosphorylate a histidine residue typically found in the histidine kinase (HK) domain. The phosphoryl group is then transferred to a conserved aspartate residue in the receiver domain of the RR, which then modulates the physiology with its C-terminal output domain often featuring a DNA-binding domain that functions as a transcriptional regulator. The amount of phosphorylated RR ultimately determines the degree of physiological modulation. In addition to the kinase activity of the SK, the concentration of phosphorylated RR may be regulated by dephosphatase activities which may occur at a number of points in a pathway. A phosphorylated receiver domain may be dephosphorylated by either the phosphatase activity of the SK or other accessory proteins (Kenney, 2010) or by the inherent catalytic activity of the receiver domain to autodephosphorylate (Hess et al., 1988, Keener and Kustu, 1988). Additionally, intermediate phosphotranfer proteins may be subjected to dephosphorylation, for example an intermediate RR Spo0F phosphotransfer protein is dephosphorylated by the RapH protein (Parashar et al., 2011). As a result, physiological modulation is imposed as a gradient over a range of activity instead of discrete on/off response. Hence, the kinase assay used to quantify the degree of phosphorylation of a

protein is the sum of kinase and phosphatase activities over a period of time under particular experimental conditions.

The core components of the TCS are the C-terminal kinase domain of the HK and the N-terminal receiver domain of the RR. There is a high degree of homology among TCSs, but the N-terminal sensor domain of the SK and the C-terminal output domain of the RR are not conserved, which reflects the diversity of the external stimuli for the SK and the types of modulation that the RR can carry out. There are two major variations of intermediate phosphotransfer proteins. In the sporulation phosphotransfer pathway of *Bacillus subtilis*, the phosphoryl group is passed down in the order of KinA, Spo0F, Spo0B, and Spo0A. KinA is the SK, and Spo0A is the RR, the final recipient of the phosphoryl group. Spo0F and Spo0B, the intermediate phosphotransfer proteins, are structurally homologous to the receiver domain of a response regulator and a HK domain, respectively (Zapf et al., 2000). The crystal structure of the complex between Spo0B and Spo0F shows that Spo0B forms a dimer like a typical HK domain, and Spo0F is associated with Spo0B with a stoichiometric ratio of 1:1. The phosphotransfer protein can be of another type, exemplified by the phosphotransfer domain of the *E. coli* ArcB domain (Kato et al., 1997). If only the fact is considered that ArcB forms a 4-helix bundle, ArcB may be regarded as being analogous to the dimerisation and histidine phosphorylation (DHp) domain of a HK domain or helical dimerisation regions of Spo0B, but key differences exist between those two classes. The DHp domain of a HK domain or helical dimerisation regions of Spo0B form symmetrical dimers with one histidine residue per monomer, but the histidine phosphotransfer domain of ArcB is a monomer consisting of a 4-helix bundle with only one histidine per structural unit (Tomomori et al., 1999). In addition to these variations of the components that can be found in a TCS system, the variation is further expanded such that some proteins form fusion proteins (Fig. 1.1).

Given a typically large number of TCS proteins in a single cell, the interaction between the SK and the RR must be specific in a given TCS pathway, otherwise the cell will not be able to adapt to the environment by responding to the stimuli properly. Structural and molecular studies have shown that the fidelity between an SK and its cognate RR is maintained via numbers of residues at the interaction interface between two proteins which co-vary to avoid cross-talk with other TCS proteins (Capra et al., 2010, Skerker et al., 2008). An exceptional case where cross-talk is present between two pathways has also been documented (Ninfa et al., 1988). CheA and NRII, the SKs that regulate the

chemotaxis and transcription of nitrogen-regulated genes, respectively, were found capable of phosphorylating the cognate RR of each other. The case is particularly interesting because NRII and CheA are Class I and II SKs, respectively (Section 1.2.2).

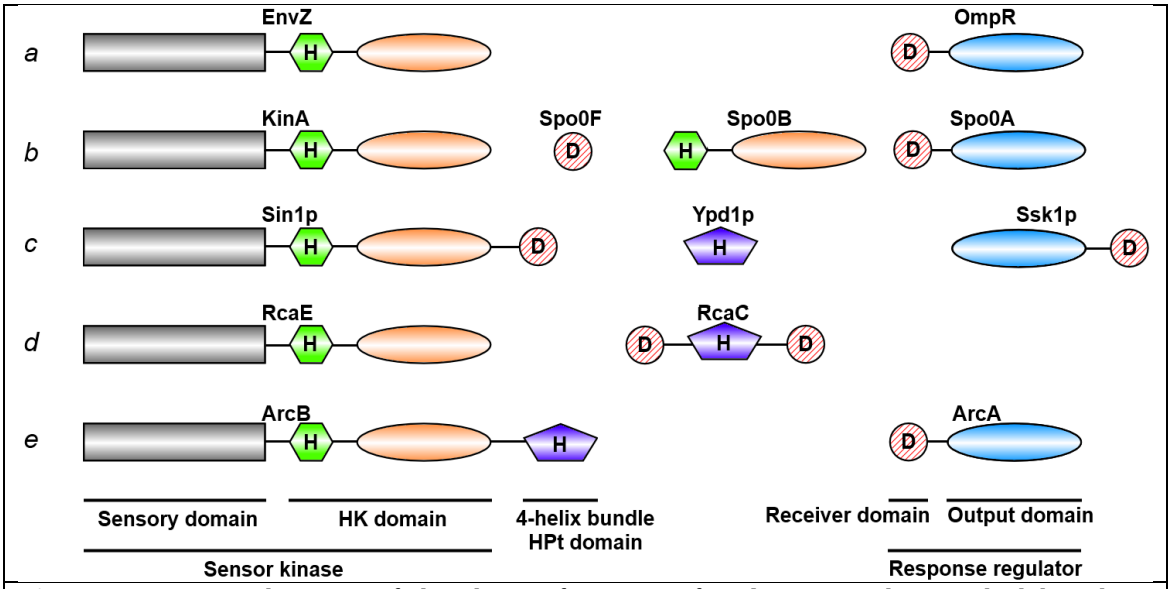


Figure 1.1 Some combinations of phosphotransfer proteins found in TCS pathways which have been documented to date. By no means are the combinations shown herein complete. Proteins are labelled and depicted schematically as monomers regardless of their biological oligomeric state *in situ*. For each given combination, an example pathway is given. H represents the conserved phosphoryl group accepting histidine residues in SK or phosphotransfer proteins, while D indicates the conserved aspartate in RR. Identical colour and shape signify the homology of the components. a, *E. coli* osmoregulation system; b, *Bacillus* subtilis sporulation pathway; c, *Saccharomyces cerevisiae* osmosensor system; d, *Fremyella diplosiphon* chromatic adaptation system; e, *E. coli* anaerobic control system. Figure adapted from Kato et al. (1997).

1.2.2 Classification of the TCS components

Diversity of each component protein involved in the TCS allows the organism to respond to a multitude of external stimuli and to elicit different types of responses. For this reason, as of present there is no universal classification method that addresses every attribute of sensor kinases. There are nevertheless numbers of ways to classify sensor kinases. These are described below.

The HK domains in SKs are identified on the basis of amino acid sequence similarities at five regions termed the N, G1, F, G2 and the H box, the last of which contains the histidine residue that is the substrate for phosphorylation (Parkinson and Kofoid, 1992). The relative positions of these elements are different among SKs, and SKs can be classified into either Class I or II accordingly (Fig. 1.2).

Class I SKs contains the H-box in the dimerisation domain adjacent to the catalytic and ATP-binding (CA) domain, hence it is referred to as dimerisation and histidine phosphorylation (DHp) domain (Song et al., 2004). Class II SKs, exemplified by the chemotaxis regulation protein CheA, have a different relative domain orientation. The most significant difference is that the H-box is not present in the dimerisation domain, but it is found in the N-terminal P1 domain, featuring a HPt domain-like histidine-containing phosphotransfer domain with a 4-helix bundles per polypeptide. The phosphoryl group is transferred from the N-terminal H-box to the response regulator via specific recognition mediated by the P2 domain.

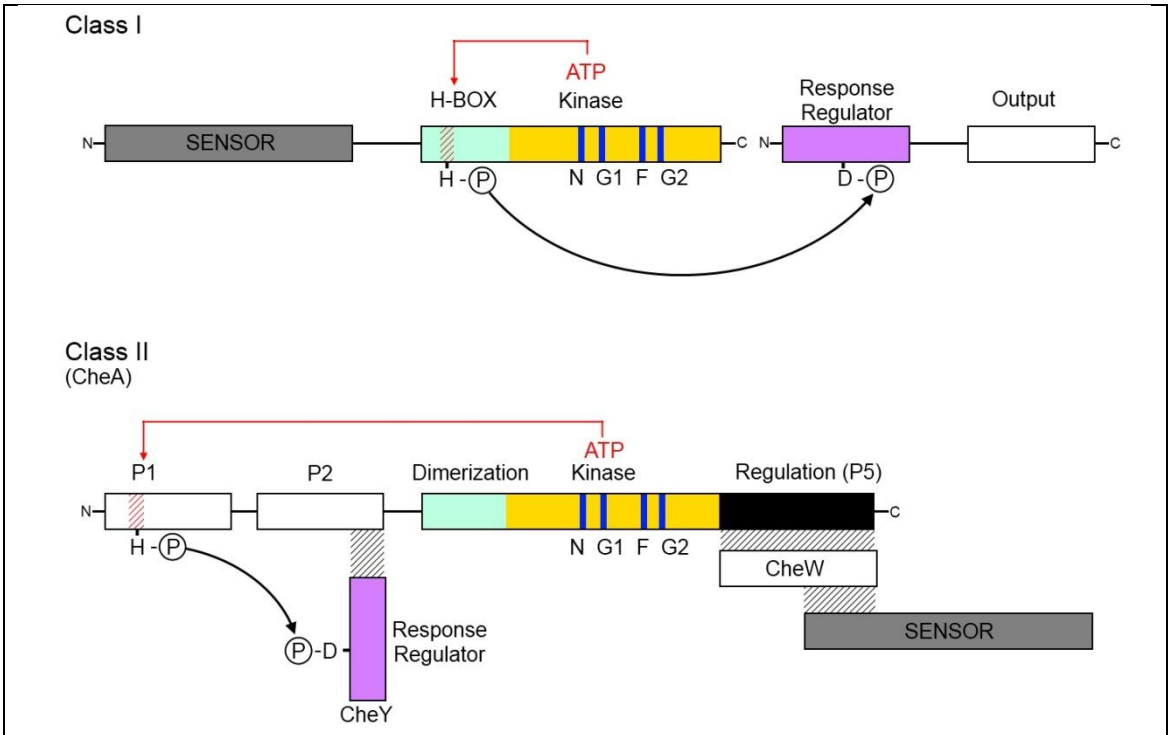


Figure 1.2 Two classes of SK protein. SK can be classified according to the position of the H-box relative to the catalytic and ATP-binding domain. Red arrows indicate the catalytic activity exerted to the histidine residue in the H-box, the substrate. Black arrows indicate the phosphotransfer from the HK domain to the response regulator. Black diagonal lines in the Class II indicate the specific interaction between the proteins involved in the system. Figure adapted from Bilwes et al. (1999).

An SK protein can either be membrane associated or found in the cytoplasm. Additionally, the position of the sensory domain relative to the transmembrane regions can vary in cases of membrane associated SKs (Cheung and Hendrickson, 2010). A prototypical SK possesses the sensor domain in between two transmembrane regions such that it protrudes into the periplasmic space. However variations of SKs include those without an obvious sensor domain, but instead the sensor domain is embedded within the transmembrane region (Fig. 1.3).

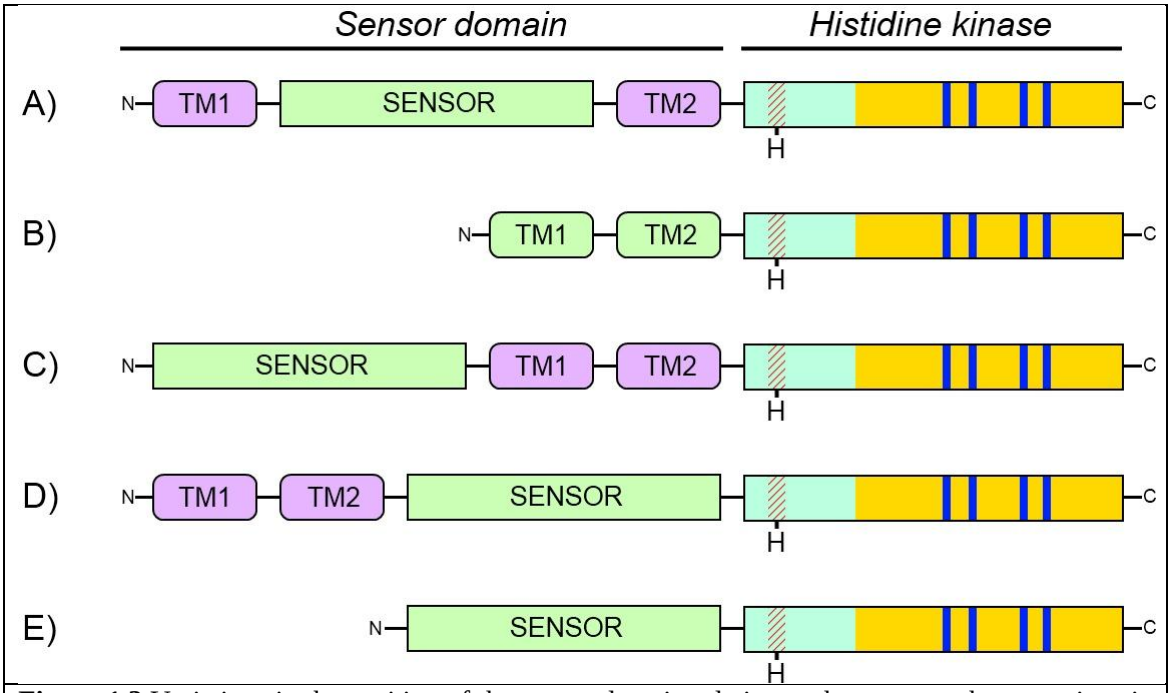


Figure 1.3 Variations in the position of the sensor domain relative to the transmembrane regions in Class I SKs. A) The sensor domain is located between two transmembrane regions, hence the former protrudes into the periplasmic space. B) The sensor domain is imbedded within the transmembrane regions. The sensor domain is N-terminal C) or C-terminal D) to the transmembrane regions, and the sensor domain is located at the cytoplasm. E) The transmembrane region is absent for a cytoplasmic histidine kinase. Transmembrane segments labelled as TM1 and TM2 may consist of several transmembrane segments, but its number will always be an odd number. Figure adapted from Cheung & Hendrickson (2010).

Moreover, detailed analyses of the sequences led to the classification of members of the SK superfamily into three superclusters which are subdivided into twelve families including a recently identified unique class of HK domain termed "HWE kinase" (Karniol and Vierstra, 2004). HWE kinases are characterised by the lack of an F-box, the presence of a histidine in the N-box and a tryptophan-x-glutamate motif in the G1-box. HWE kinases are found in α- and γ-proteobacteria, but concentrated in the *Rhizobiaceae* family. HWE kinases are often found as a fusion protein with the response regulator attached to the C-terminus of the HK domain.

In addition to the above-mentioned diversity of SK, some signalling proteins with SK ancestry do not function as a HK. Examples include chloroplast sensor kinases (CSK, Puthiyaveetil et al., 2008), ethylene receptors (Moussatche and Klee, 2004), and plant phytochromes (Chen et al., 2004). In the latter two, the HK-like domains have evolved to function as Serine/Threonine kinases.

Detailed account of the modular nature of the signalling proteins other than SKs in bacteria lies outside the scope of this introduction. Nonetheless, an example where

diversity is found in the C-terminal signalling domain relative to the conserved N-terminal sensor domain is briefly described here. Six bacteriophytochrome genes are present in the genome of *Rhodopseudomonas palustris* CGA009 (Giraud et al., 2005, Larimer et al., 2004). These encode proteins of which all but RpBphP1 have a histidine kinase at the C-terminus (Fig. 1.4). The C-terminal output transducing domain (OTD) of RpBphP1 comprises a PAS/PAC domain followed by a 2-helix output sensor (HOS) domain (Bellini and Papiz, 2012b). This observation highlights the plasticity of the bacterial genomes and their ability to carry out horizontal gene transfer (Dobrindt and Hacker, 2001).

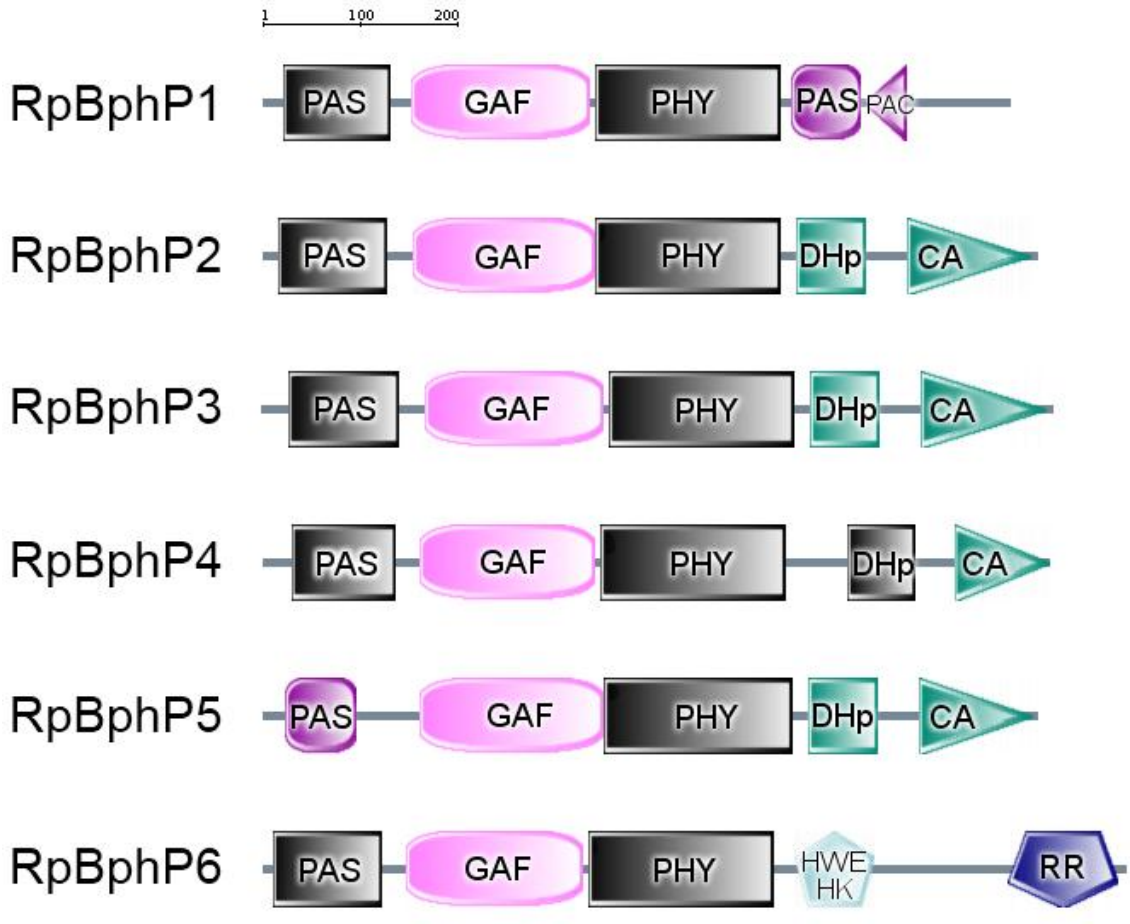


Figure 1.4 Homologues of bacteriophytochromes found in the genome of *Rhodopseudomonas palustris* CGA009. The translated coding sequences were submitted to the SMART domain prediction server with the PFAM option enabled. The accession code of the genome sequence used is NC_005296 (Larimer et al., 2004), and the gene locus for each protein is as follows; RpBphP1 – rpa1537 (frame shift mutation corrected by inserting a guanine base at 430th position, Braatsch et al., 2007), RpBphP2 – rpa3015, RpBphP3 – rpa3016, RpBphP4 – rpa1490, RpBphP5 – rpa0122, RpBphP6 - rpa0990. Accession numbers are as mentioned in Rockwell et al. (2006). All but RpBphP1 possess a histidine kinase with canonical domain organisation except for RpBphP6 which is a fusion protein typical for a HWE histidine kinase. RpBphP1 has no histidine kinase, and illustrates the fact that the modularity is not limited to histidine kinases, but is widespread among many bacterial signalling proteins. The scale shown at the top indicating the number of amino acid residues applies to all the schematic diagrams.

1.2.3 Structures of sensor kinases

A wealth of structural information for TCSs has now accumulated. Structurally analysed histidine kinases and phosphotransfer proteins are listed in Table 1.1.

A canonical Class I histidine kinase consists of a DHp domain and a CA domain (Fig. 1.5), and must form a parallel dimer to function as a kinase. A DHp domain consisting of two α-helices forms a four-helix bundle upon dimerisation (Tomomori et al., 1999). The CA domain is a member of the GHKL (GyraseB, Hsp90, histidine kinase and MutL) superfamily (Dutta and Inouye, 2000). The members of this family assume a common ATP-binding Bergerat fold, the core of which minimally consists of a mixed α/β sandwich with 4 anti-parallel β-strands and 3 α-helices. In addition to the variable N-terminal sensor domain and the conserved C-terminal histidine kinase domain, a HAMP (Histidine kinase, adenylyl cyclase, MCP and phosphatase) linker may be found between the two. A HAMP domain is found in over a fifth of SKs and two-thirds of chemotaxis receptors. It consists of two parallel α-helices connected by a long linker, and these two parallel helices are off-set in the longitudinal direction by a single turn of an α-helix. The HAMP domain dimerises in solution to form four-helix bundles (Fig. 1.5, Hulko et al., 2006).

Source organism	Protein	PDB accession code	Reference
Histidine kinase domain / HK+RR complex			
Thermotoga maritima	CheA	1B3Q	Bilwes et al. (1999)
Thermotoga maritima	HK853	2C2A	Marina et al. (2005)
Geobacillus stearothermophilus	KinB with Sda	3D36	Bick et al. (2009)
Thermotoga maritima	HK853 and RR468	3DGE (3DGF, 3GL9)	Casino et al. (2009)
Thermotoga maritima	ThkA with TrrA	3A0R	Yamada et al. (2009)
Escherichia coli	QseC	3JZ3	Xie et al. (2010)
Dimerisation domain			
Escherichia coli	EnvZ	1JOY	Tomomori et al. (1999)
Escherichia coli	CheA	Preliminary study	Park et al. (2011)
Catalytic and ATP-binding domain			
Escherichia coli	EnvZ	1BXD	Tanaka et al. (1998)
Escherichia coli	PhoQ	1IDO	Marina et al. (2001)
Thermotoga maritima	CheA	1I58, 1I59, 1I5A, 1I5B, 1I5C, 1I5D	Bilwes et al. (2001)
Escherichia coli	NRII (NtrB)	1R62	Song et al. (2004)
Salmonella typhimurium	PhoQ	3CGZ, 3CGY	Guarnieri et al. (2008)
Aquifex aeolicus VF5	HksPr	Preliminary study	Horita et al. (2011)
	PrrB	** E. Nowak, S. Panjikar & P. Tucker, unpublished work	
Phosphotransfer protein complex			
Bacillus subtilis	Spo0F and Spo0B	1F51	Zapf et al. (2000)
Saccharomyces cerevisiae	YPD1 and SLN1-R1	1OXB, 1OXK	Xu et al. (2003)
Bacillus subtilis	Spo0F and Spo0B	2FTK	Varughese et al. (2006)
Saccharomyces cerevisiae	YPD1 and SLN1-R1	2R25	Zhao et al. (2008)
Bacillus subtilis	Spo0F and RapH	3Q15	Parashar et al. (2011)
Phosphotransfer protein			
Escherichia coli	ArcB	1A0B	Kato et al. (1997)
Bacillus subtilis	Spo0B	1IXM	Varughese et al. (1998)
Saccharomyces cerevisiae	YPD1	1C02, 1C03	Song et al. (1999)
Saccharomyces cerevisiae	YPD1	1QSP	Xu, Porter and West (1999)
Salmonella typhimurium	CheA	1I5N	Mourey et al. (2001)
Thermotoga maritima	CheA	2LD6	Vu et al. (2011)
Miscellaneous			
Thermotoga maritima	CheW	Preliminary study	Park and Crane (2011)

Table 1.1 Information about structural studies carried out on TCS component proteins to date except for RRs. ** Unpublished work on PrrB was mentioned in Horita et al. (2011).

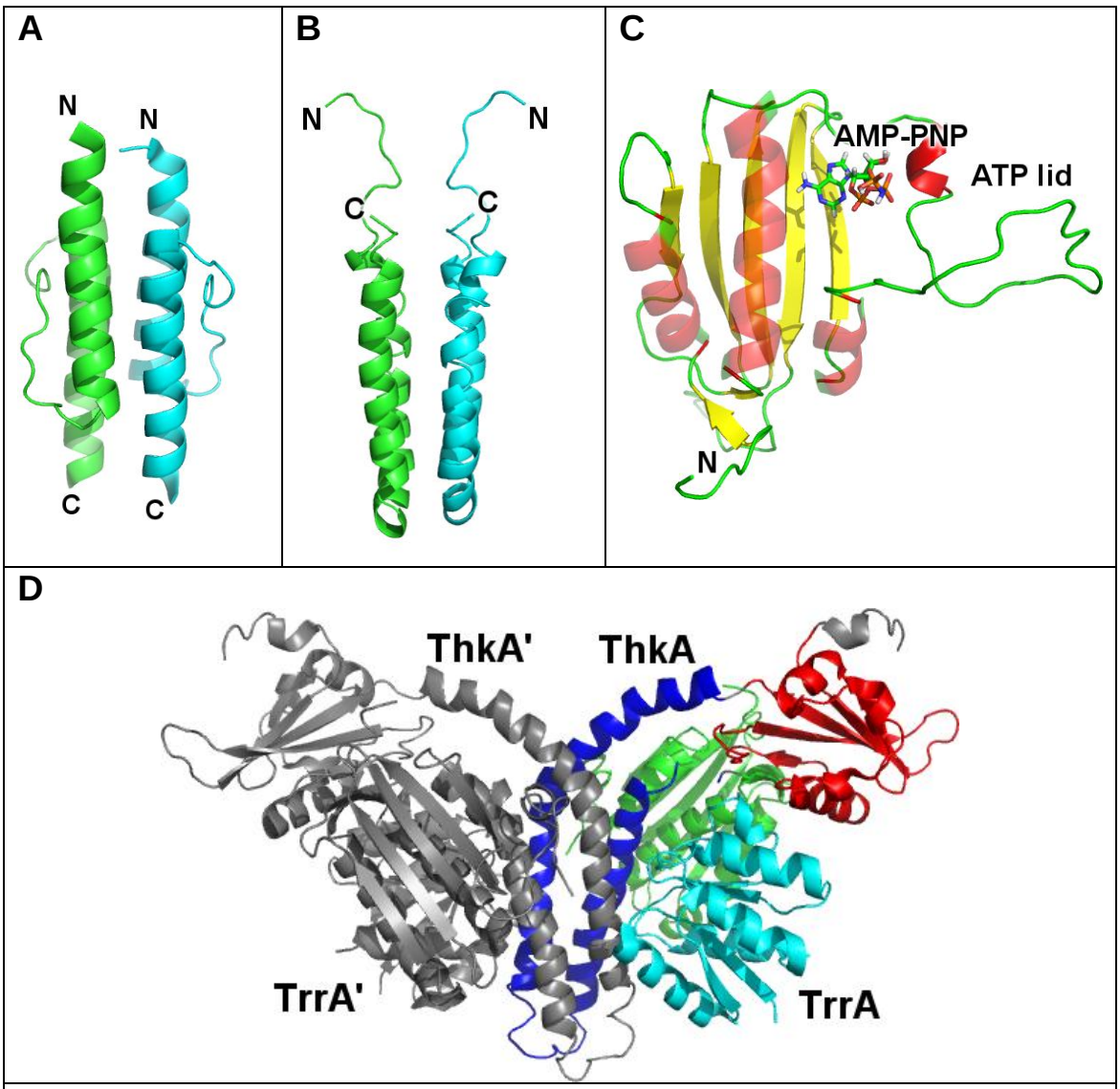


Figure 1.5 Domains of histidine kinase. A) HAMP domain of Af1503 of *Archaeoglobus fulgidus* shown as a dimer with each monomer coloured differently. PDB accession code: 1ASX (Hulko et al., 2006). B) DHp domain of EnvZ of *E. coli* shown as a dimer with each monomer coloured differently. PDB accession code: 1JOY (Tomomori et al., 1999). C) CA domain of EnvZ. Three α-helices that are conserved among the GHKL superfamily are presented transparently to aid the visibility of β-strands. PDB accession code: 1BXD (Tanaka et al., 1998). The 1BXD structure was solved as a complex with AMP-PNP, a non-hydrolysable structural analogue of ATP. The ATP lid in this structure is thought to correspond to the open conformation of the proteins of the GHKL superfamily (Dutta and Inouye, 2000). D) PAS–HK (ThkA) and RR (TrrA) complex of *Thermotoga maritima*. The domains of a ThkA monomer are coloured as follows: PAS (red), DHp (blue), and CA (green). One TrrA monomer is coloured in cyan. The symmetry related ThkA and TrrA molecules are coloured grey. PDB accession code: 3A0R (Yamada et al., 2009).

1.2.4 Mechanism of signal transduction

To date the mechanism by which the external stimulus, perceived by the sensor domain, is translated into modulation of the kinase activity and the extent of conservation of the mechanism is not fully understood. The motivations for understanding the mechanism of signal transduction are manifold, of which two major ones are as follows. Firstly, TCSs as potential targets of antimicrobial drugs have previously been discussed (Matsushita and Janda, 2002). A TCS functions to sense numerous external stimuli and to elicit responses, some of which are involved in controlling the virulence in bacteria, a notable example being quorum sensing. Means by which to intervene with signal transduction may render pathogenic bacteria unable to coordinate their pathogenic activities. Since a TCS is not found in mammals, potential side effects from a compound targeted to a TCS are expected to be minimal. Secondly, the modular nature of SKs has enabled production of functional chimera as described later. Characterising the mechanism of signal transduction will enable a more rational approach in designing a component of a synthetic biological circuitry.

As previously described, an SK must form a dimer to be catalytically active. The notion that dimerisation of an SK is an absolute prerequisite for its functionalities is in accordance with observations that the HK domains, HAMP domains, and some sensor domains form dimers, as well as with the proposed mechanism of autophosphorylation occurring in *trans*, i.e. monomers in a dimer phosphorylate each other. However, recent evidence has shown that at least some SKs undergo intramolecular phosphorylation by a *cis* mechanism (intra-molecular autophosphorylation), which still requires dimerisation (Casino et al., 2009).

Regarding the mechanism of signal transduction along transmembrane α-helices, the following four models have been proposed: A) Horizontal translation along the plane of the membrane, B) piston movement along the axis normal to the membrane, C) pivot movement along the axis parallel to the membrane, and D) rotation along the axis perpendicular to the membrane (Matthews et al., 2006). An adaptation of these models to SKs with a transmembrane α-helix is depicted in Fig. 1.6. These models of signal transduction were proposed for transmembrane α-helices, but they do not necessarily exclude soluble SKs. According to those models, the signal is transduced by movements of monomers relative to each other. Alternatively, Möglich and co-workers propose another model whereby the signal transduction takes place by the rotation of the linker

between the sensory and the signalling domain, and the signal is transduced by the change in the relative position of two domains (2009). To date a number of functional chimeric SKs have been generated that may suggest that the mechanism of inter-domain signal transduction is conserved, not just within the closely related SKs, but also within diverse signalling proteins.

The conservation of a signal transduction mechanism could be anticipated for closely related SKs with similar domain organisations and, indeed, functional chimeric SKs have been created by fusing similar parental SK domains (Baumgartner et al., 1994, Feng et al., 1997, Krikos et al., 1985, Perraud et al., 1998, Utsumi et al., 1989, Ward et al., 2002, Weerasuriya et al., 1998, Zhu and Inouye, 2003). Experimental evidence for a signal transduction mechanism was lacking until 2006 when the structure of a HAMP domain was solved (Hulko et al., 2006). Based on the coiled-coil structure of 4-helix bundles with unusual "knobs-to-knobs" packing (rather than the usual "knobs-into-holes" packing), helix rotation was proposed to occur as a result of switching between two isoenergetic packing modes, rather like cogwheels. Evidence in agreement with the rotation model came from a study utilising a chimeric SK where insertions and deletions were introduced at the linker region between the sensor domain and the HK domain. The length of the linker region affects how the kinase activity is modulated by the sensory domain. Moreover, the stimuli-response pattern exhibited a heptad periodicity as insertions/deletions were introduced to the linker (Möglich et al., 2009). Data in partial agreement with the above-mentioned rotation model was presented in a study with Tez1, a chimeric protein with the periplasmic and transmembrane regions of Tar1 fused to the HAMP domain and the HK domain of EnvZ (Zhu and Inouye, 2003). Therein, the number of alanine residues inserted at the HAMP domain affected the ability of Tez1 to modulate the kinase activity in response to the stimulus. Inasmuch as the stimuli-response pattern was affected by the number of residues in the linker region, this observation is in accordance with the rotation model. However in the same study, the type of mutation introduced in the HAMP domain also had a profound effect on the kinase response to the stimulus, implying that it is not just the length (which affects the rotational orientation of the HK domain), but also the properties of the linker region that play a key role in signal transduction. It must be noted that the study was carried out prior to the structural information of the HAMP became available, and some of the mutations resulted in disruption of the heptad periodicity of hydrophobic residues that is

proposed to be essential in the rotation model (Hulko et al., 2006). Therefore the rotation model is not entirely refuted by Zhu and Inouye (2003).

Examples of functional chimeras created from more diverse proteins include a fusion of the sensor domain of a soluble SK (*Synechocystis* PCC6803 phytochrome 1, Cph1) and the HAMP domain and the HK domain of *E. coli* EnvZ, a membrane associated SK (Levskaya et al., 2005) and a fusion of the LOV domain of YtvA and the HK domain of FixL (Möglich et al., 2009). The notion that these diverse proteins function by the same signal transduction mechanism is still plausible given the variation of modular proteins observed in nature. The phytochrome homologues RpBphP1 and RpBphP2 for example possess different output domains despite having similar sensor domains (Fig. 1.4). One the other hand, a wide variety of SKs share common HK domains despite their diverse N-terminal sensor domains.

A functional chimera termed "AIR" was created by fusing the periplasmic and transmembrane regions of *E. coli* Tar (referred to as aspartate receptor, AR therein) and the human insulin receptor (IR). The signal transduction mechanism was suggested to be conserved even between bacterial and human receptor proteins (Moe et al., 1989). The original assumption was that AR and IR share a similar topology and may therefore operate by a similar mechanism. This study must be approached carefully since it was carried out prior to the detailed structural determination of the human IR becoming available (Lawrence et al., 2007) which demonstrated gross architectural differences between human IR and bacterial signalling proteins. As far as is known, this controversial hypothesis of a universally conserved signal transduction mechanism has not been tested experimentally to a degree comparable with the "rotation model" for the SK by Hulko et al. (2006) and Möglich et al. (2009). Another chimera termed "IAR" features an alternative combination between IR and AR, and it failed to exhibit kinase activity in response to the stimulus (Ellis et al., 1986). The apparently functional "AIR" chimera may therefore be an artefact and it alone is insufficient to prove the notion of a universal mechanism for receptor proteins.

It is clear that a conserved mechanism exists between various receptor proteins to a certain extent, but the exact boundary of conservation of the mechanism is currently unknown. Questions remain over the appropriateness of the "rotation model" for some SKs whose sensor domain fragments assume the monomeric form (Cheung and Hendrickson, 2010) or signalling proteins with an atypical domain arrangement. If the

sensor domains of these proteins do not form a molecular contact in a dimer, the rotational force will not be transferred to the signalling domain, because the rotational force generated by the sensor domain will result in its own rotation relative to the HK. The soluble SK RpBphP4 could correspond to such a hypothetical case. RpBphP4 was shown to assume a Y shaped global conformation in solution where the N-terminal sensor domains were interpreted as being far apart (Evans et al., 2006). Af1503 of *Archaeoglobus fulgidus* exemplifies another unusual case because the HAMP domain is its sole cytoplasmic part (Hulko et al., 2006). It seems unlikely for these unusual proteins to operate by the same mechanism as SKs with a canonical domain organisation.

The mechanism of signal transduction by helical rotation seems to be widespread, but other models of signal transduction cannot be excluded. Whether other models also play roles in the signal transduction in SKs, and if so, how widespread such mechanisms are is yet to be seen.

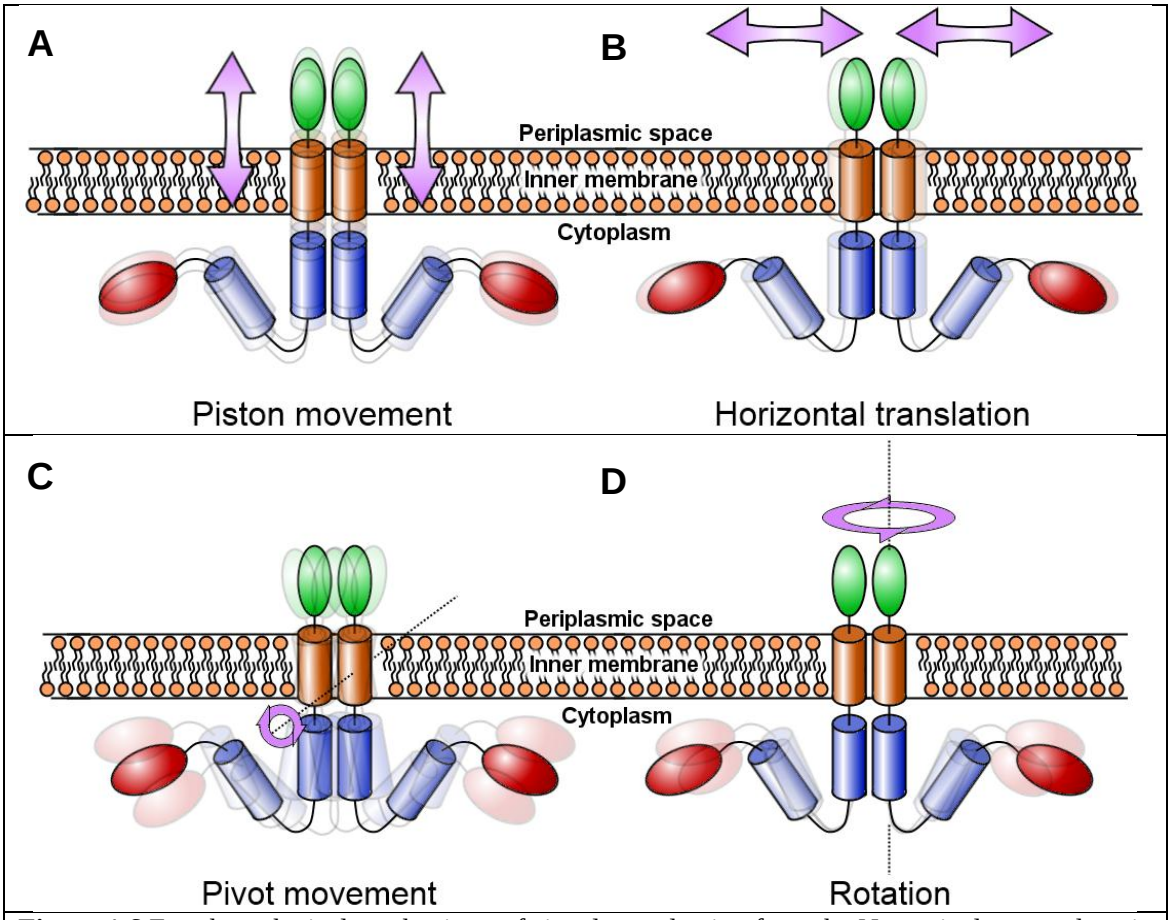


Figure 1.6 Four hypothetical mechanisms of signal transduction from the N-terminal sensor domain to the C-terminal HK domain in a membrane associated Class I SK. A) Piston movement, B) Horizontal translation, C) Pivot movement, and D) Rotation model. The colour scheme is as follows Periplasmic sensor domain: Green, transmembrane region: Orange, DHp domain: Blue, and CA domain: Red. These motions will cause movements relative to each monomer.

1.3 Phytochrome

1.3.1 Diversity and evolutionary perspectives

Phytochrome was originally identified in plants as being a light sensitive pigment that absorbs light most effectively in the red region (Butler et al., 1959). Phytochromes are light-sensitive proteins that reversibly interconvert between two thermostable forms termed Pr-form, the red-light absorbing form, and Pfr-form, the far-red light absorbing form. Phytochromes regulate numerous light dependent physiological responses (Scheerer et al., 2010). Since the original discovery, phytochromes have been identified in cyanobacteria, photosynthetic bacteria, non-photosynthetic bacteria, filamentous fungi, slim moulds, and diatoms (Lamparter et al., 1997, Jiang et al., 1999, Davis, 1999, Blumenstein et al., 2005, and as reviewed in Vierstra and Zhang, 2011). Additionally, variants of phytochrome with unusual domain organisation (Ulijasz and Vierstra, 2011) and photoreceptors that are related to phytochrome, but with a distinctly different photocycle, have been identified (Rockwell et al., 2011).

The diversity of phytochromes has led to at least two hypotheses being put forward regarding their evolutionary path. The plant phytochrome was originally thought to have been inherited from cyanobacteria during the process of endosymbiosis. The theory is based on similarities between the plant and the cyanobacterial phytochromes in their cognate bilin chromophore and the location of the covalent bond forming cysteine residue (Herdman et al., 2000, Lamparter et al., 2003). An alternative theory hypothesises that the plant and the cyanobacterial phytochromes are distantly related to each other based on sequence alignments carried out independently on the N-terminal and the C-terminal parts of phytochromes (Lamparter, 2004). This result led to a proposition that plant cells already possessed bacteria-type phytochromes prior to the acquisition of plastids. The endosymbiosis resulted in the introduction of biliverdin reductase from the cyanobacteria to the plant, which explains the similarity between their cognate bilin chromophore. The reduction of the bilin changed the preferred site for the covalent bond-forming cysteine residue, and the seemingly conserved cysteine residue in plant and cyanobacterial phytochromes is therefore the result of convergent evolution. Neither of these hypotheses have become the consensus, but the presence of bacterial-type phytochromes in both cyanobacteria and plants may be interpreted in favour of the latter hypothesis (Quest and Gärtner, 2004).

1.3.2 Bilin and the mode of attachment

Phytochrome's sensitivity to light originates from a covalently bound chromophore, which is a bilin, a linear tetrapyrrole. The bilin is covalently attached to a conserved cysteine residue, and the process of autocatalytic attachment is a characteristic feature of *bona fide* phytochrome, which is unusual among phycobiliproteins (Scheer and Zhao, 2008). The type of the covalently bound bilin chromophore differs among the lineages of phytochromes. In general plant phytochromes bind to phytochromobilin (PΦB), cyanobacterial phytochromes bind to phycocyanobilin (PCB), and bacterial and fungal phytochromes bind to biliverdin (BV). BV is the immediate breakdown product of haem by haem oxygenase (Lamparter, 2004). BV binds to the cysteine residue in the N-terminal region upstream of the PAS domain via its vinyl side chain (Lamparter et al., 2003). In case of plant and cyanobacterial phytochromes, BV is reduced to produce PΦB and PCB in plant and cyanobacteria, respectively. These two bilins have an ethylidene side chain at the A-ring in common, and covalently bind to the cysteine in the GAF domain (Lamparter, 2004).

The GAF domain is the minimum lyase domain required for the formation of the covalent bond between the cysteine residue and the bilin in plant and cyanobacterial phytochromes (Wu and Lagarias, 2000). For bacteriophytochrome that binds to a BV via its N-terminal region, the PAS domain is required for the covalent bond to form between the BV and the cysteine residue. Bacteriophytochromes have been shown to bind to a PCB non-covalently, but the covalent bond will form if a cysteine is present in the GAF domain at the position homologous to the plant/cyanobacterial phytochromes (Lamparter et al., 2002, Wagner et al., 2007). Whether an isolated GAF domain of a bacteriophytochrome with a cysteine introduced at the homologous location forms a covalent bond with PCB has yet to be shown.

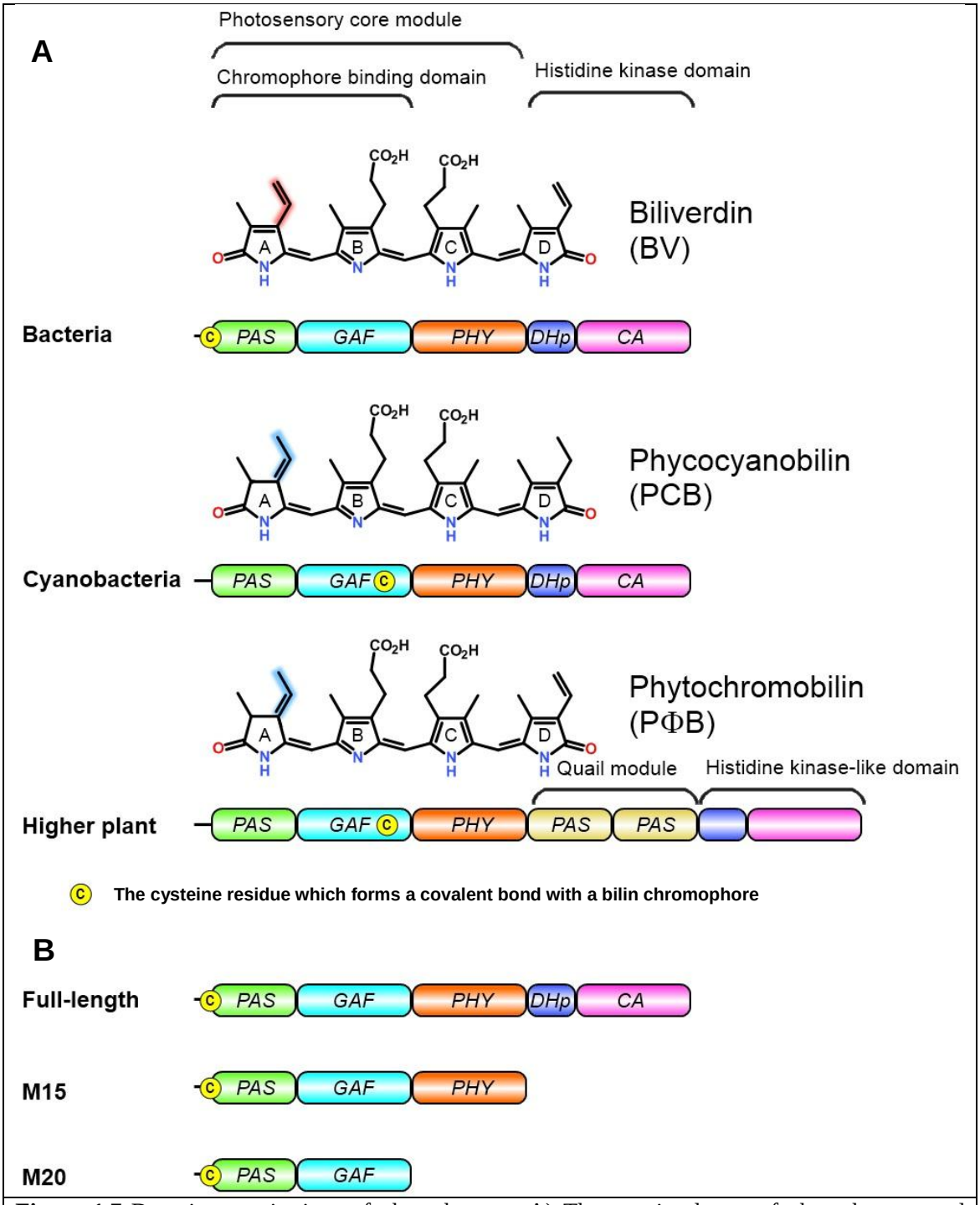


Figure 1.7 Domain organisations of phytochromes. A) Three main classes of phytochromes and their cognate bilin chromophores. Homologous domains are coloured the same. The bilins are shown in their free form. The bilin side chains that form covalent bonds with conserved cysteine residues are highlighted with colours - red for vinyl side chain of BV, and blue for ethylidene side chains of PCB and PΦB. The cysteine residues where the covalent bonds form are indicated in yellow. B) Fragments of the Agp1 phytochrome used in the present study.

1.3.3 Photochromicity

Binding of a bilin, including the formation of a covalent linkage, to its cognate apo-phytochrome causes a change in the absorption spectrum characterised by an increase in the extinction coefficient and the λ_{max} in the Q-band region. Similar changes are also observed for non-covalent adduct formation (Lamparter et al., 2002). Formation of the covalent bond also has an effect on the absorption spectrum, as the non-covalent Cph1-PCB adduct is red-shifted with respect to the covalently bound adduct (Borucki et al., 2003). The λ_{max} (Q-band) of holoprotein adducts of plant and cyanobacterial phytochromes are, in general, roughly 50 nm lower than those of bacterial phytochromes (Lamparter, 2004). An increase in absorption in the Q-band region enables the assignment of "Specific Absorbance Ratio (SAR)" (Scheerer et al., 2006), which is calculated according to the equation below.

$$SAR = Absorbance_{Qmax} \div Absorbance_{280nm}$$

Here $Absorbance_{Qmax}$ is the absorbance at the λ_{max} in the Q-band region. The SAR value enables the evaluation of the proportion of holoprotein in a given sample relative to other samples of the same protein.

Absorption spectroscopy and resonance Raman spectroscopy allowed for the identification of several intermediate stages of the photocycle of phytochromes (Fig. 1.8). In the Pr-form all the pyrrole ring nitrogens are protonated. Upon absorption of a photon, isomerisation occurs around a double bond between the C and the D rings of the bilin, leading to the formation of the Lumi-R form within a picosecond timescale. Lumi-R (λ_{max} = 688 nm), with a partially bleached absorption spectrum, then undergoes thermal relaxation into Meta-Ra, with λ_{max} at 663 nm, on a microsecond timescale. The subsequent formation of Meta-Rc (λ_{max} = 725 nm) on a micro- to millisecond timescale is associated with deprotonation of a pyrrole ring nitrogen. Finally, the thermal relaxation steps are completed with the formation of the Pfr-form with fully protonated pyrrole ring nitrogens. These relaxation stages induce conformational changes in the protein that ultimately result in modulation of the activity of the C-terminal kinase domain. The Pfr-form can revert back into the Pr-form either by a spontaneous process of dark reversion or by absorbing far-red light, the latter proceeding at the faster rate. Some phytochromes referred to as bathy-phytochromes (Karniol and Vierstra, 2003) will undergo dark reversion in the opposite direction such that the Pfr-form is the dark-adapted form. Neither Pfr→Pr pathways are well understood (Wagner et al., 2008). To

date, photocycles of the classical (stable in the Pr-form) and the bathy-phytochromes (stable in the Pfr-form) have not been fully established as proceeding by the same pathways, although some similarities including the initial D-ring flip have been reported (Yang et al., 2011, and as reviewed in Scheerer et al., 2010).

The changes that occur in the bilin chromophore during photoconversion have been a subject of intense investigation by a multitude of techniques. The current consensus for the first stage is that the D-ring flips via an isomerisation around the double bond between the C and the D-rings (Rockwell and Lagarias, 2010, Song et al., 2011, Yang et al., 2011). The first crystallographic evidence for this came from crystal structures of the PaBphP Q188L mutant. Crystal structures of a mutant with markedly decreased rate of dark-reversion were deemed to be in a mixed state of a Pfr/Pr-form as judged by absorption spectroscopy (Yang et al., 2011). The electron densities of the bilin chromophore and the surrounding residues were broad and this was interpreted to be the result of the simultaneous presence of BV in two different configurations – one, as in the wild type PaBphP crystal structure, which is in a pure Pfr-form (Yang et al., 2008) and the other, as in the DrBphP crystal structure, which is in a pure Pr-form (Wagner et al., 2007). More explicit evidence was obtained from a "trap–pump–trap–probe" experiment carried out on PaBphP, where the data suggest that the configurational change is initially confined in the D-ring, but later the changes extends to the B- and then to the A-ring (Yang et al., 2011). Collectively, the data supports the "flip-and-rotate" model proposed for the overall movement of the bilin during the photoconversion (Yang et al., 2009). The initial D-ring flip is followed by a clockwise in-plane rotation of the entire bilin moiety relative to the GAF domain (view direction onto the α-face of the bilin). With regards to the initial D-ring flip, the directions of the rotation of the D-ring in the plant/cyanobacterial and bacterial phytochromes were proposed to be the opposite. The D-ring of the BV in bacteriophytochromes and the D-ring of the phytobilin (PΦB and PCB) in plant/cyanobacterial phytochromes were proposed to rotate anti-clockwise and clockwise, respectively, as the bilin is viewed in parallel to the planar B- and C-rings with the D-ring in the foreground (Rockwell et al., 2009).

A surprising observation has been reported where the primary movement of the bilin occurs at the A-ring in the GAF domain fragment of *SyB*-Cph1 from *Synechococcus* OS-B'. Comparison of the solution structures of *SyB*-Cph1 in the Pr- (PDB code 2KOI)

and the Pfr-form (PDB code 2KLI) showed the rotation to occur at the A-ring around the axis of C4=C5 atoms by about 90°. The bilin structures were significantly remodelled in the refined structures of *SyB*-Cph1 in the Pr- (2LB9) and the Pfr-form (2LB5). Changes include the positions of the bilin moieties relative to the GAF domains in both forms. Additionally, the movement of the A-ring as inferred from the refined structures in the Pr- and the Pfr-forms is qualitatively different from the original proposal. Nonetheless the conclusion that the A-ring is the major location of the conformational change during the photoconversion remains the most favourable outcome from the comparison of the revised structures. This finding was commented as controversial and difficult to reconcile with the existing data (Hughes, 2010), but given the recent discovery of more diverse members of the phytochrome family, some of which undergo complex photocycles involving two cysteine residues (Rockwell et al., 2011), it could be hypothesised that the molecular trajectories during the photoconversion may not be conserved among all classes of phytochrome and related proteins. Additionally, there is always a concern when comparing structures solved by crystallography and by nuclear magnetic resonance (NMR) spectroscopy. Indeed, effects of crystal packing on the global conformation have been described previously for Cph1 (Mailliet et al., 2011) and are as well described for Agp1 in the present study (Section 3.10.2). It is possible that the process of crystallisation led to accumulation of structurally homogeneous species of phytochrome molecules whereas data from NMR spectroscopy may be reflecting the simultaneous presence of more diverse species at different points in the photocycle. It will be interesting to obtain the crystal structures of *SyB*-Cph1 and compare the solution and crystal structures for any differences between them.

Configurations of the bilins in phytochrome holoproteins are difficult to study, because of their movement during the photocycle under ambient condition. To this end, synthetic, locked chromophores with additional carbon chain(s) that maintain the predetermined configurations were devised (Figure 1.9). The holoprotein adducts with locked chromophores were analysed for their resemblance to the natural holoproteins. UV-vis spectra collected from phytochrome adducts with singly locked chromophores termed 15Za and 15Ea strongly suggest that the structures of these synthetic molecules to resemble BV in the Pr- and the Pfr-forms, respectively in both Agp1 and Agp2, and these are consistent with the configurations of the BVs found in the crystal structures of DrBphP (5Zs/10Zs/15Za, Wagner et al., 2007) and PaBphP (5Zs/10Zs/15Ea, Yang et al.,

2008) . Analyses on Agp1 and Agp2 holoprotein adducts with the locked chromophores 5Za and 5Zs indicated that the holoproteins with 5Zs have absorption spectrum that more closely resembles that of the natural holoprotein in the Pr-form. However, the 5Za holoprotein was found to resemble the Pfr-form more closely than the 5Zs holoprotein upon photoconversion, thus indicating the C5-C6 single bond to be in the *anti* conformation in the Pfr-form. Motion around the A-ring in the photocycle is consistent with the mechanism proposed for *SyB*-Cph1 (Ulijasz et al., 2010). However, the *anti* conformation around the C5-C6 bond is inconsistent with the *syn* conformation suggested for the Pfr-form of Cph1 (Hahn et al., 2008, Rohmer et al., 2008) and found in the PaBphP crystal structure (Yang et al., 2008). A relatively broad electron density calculated from a low resolution dataset may not exclude the possibly of an alternative configuration of the BV in the structure in the latter case.

Additionally, the identity of the photocycle between the classical and the bathy-phytochrome has not been fully established. A notion that the C5-C6 bond assumes the *anti* conformation is further supported by a study employing four doubly locked chromophores in a study to determine the stereochemistry between the A- and the B-rings (Inomata et al., 2009). The absorption spectra of Agp1 and Agp2 holoproteins with the doubly locked chromophores in the 5-*anti* conformation resemble those of the natural holoproteins, but those of chromophores with the 5-*syn* conformation do not. Out of two doubly locked chromophores with the 5-*anti* conformation, the Agp2 holoprotein with 5Ea15Ea conclusively resembled the Pfr-form more than did the holoprotein with 5Za15Ea, but the resemblance was not so clear in case of Agp1. Because of the ambiguity concerning the correct stereochemistry between the A- and the B-rings it seems possible that the stereochemistry of the bilins is different among phytochromes. In a recently study, synthetic locked chromophores termed 15ZaPCB and 15EaPCB were used, and those PCB chromophores resemble the configurations of 15Za and 15Ea, respectively (Yang et al., 2012). The 15EaPCB adduct of plant phytochromes exhibited Pfr-like properties during *in vitro* and *in vivo* studies, thus supporting the notion of conserved chromophore motion for phytochromes with canonical domain organisation.

Clearly more research is required to establish the exact molecular trajectory of stereochemical changes at the bilin in phytochrome during the photocycle and also the extent of conservation of the mechanism in the phytochrome family.

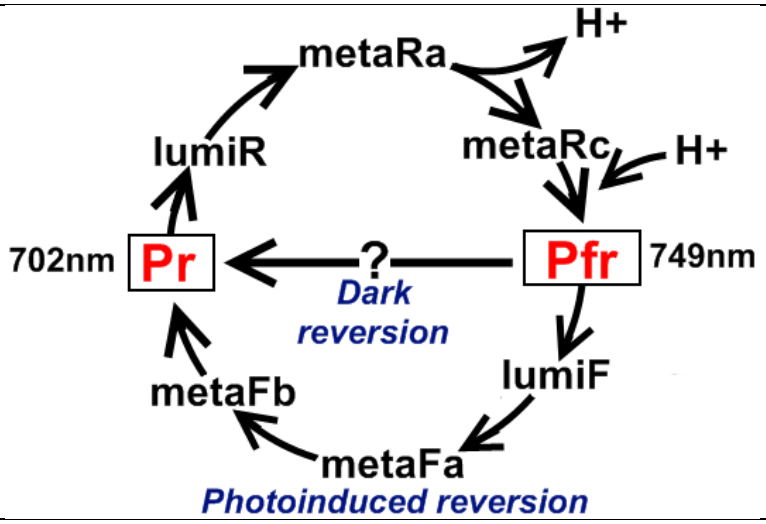


Figure 1.8 Schematic diagram of the photocycle of a classical phytochrome. Diagram adapted from Wagner *et al*., 2008. The wavelength of the absorption maximum in the Q-band of the Pr- and the Pfr-forms are of Agp1 M15-BV.

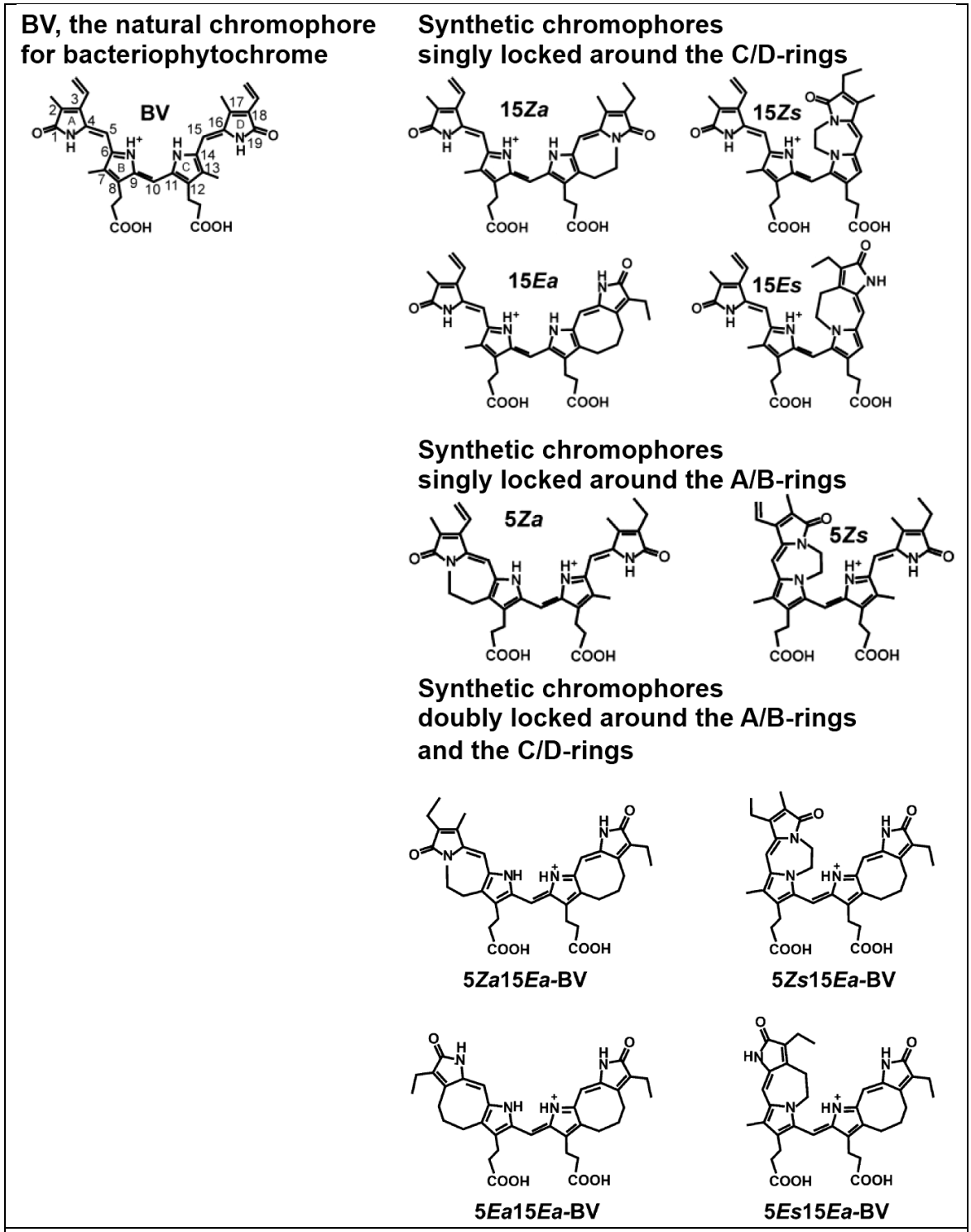


Figure 1.9 All locked chromophores that were tested on bacteriophytochromes to date. The numbers in their names indicate the number of the carbon atom where the single and the double bonds have been locked in a certain stereochemistry. *Z* and *E* represent the configuration of the double bonds, respectively, whereas *s* and a represent *syn* and *anti* conformations of single bonds, respectively. Note all the doubly locked chromophores feature ethyl (instead of vinyl) side chains, therefore would not allow the formation of covalent adducts with bacteriophytochromes. Figures adapted from Inomata et al. (2005); Inomata et al. (2006); Inomata et al. (2009).

1.3.4 Phytochrome structures

Phytochrome domain organisation has been known based on sequence alignment prior to any structure determination (Montgomery and Lagarias, 2002). The N-terminal sensor domain of phytochromes, referred to as the photosensory core module (PCM, Yang et al., 2009), consists of a PAS, a GAF, and a PHY domain (Fig. 1.7), all of which were found to possess a conserved fold (Essen et al., 2008, Yang et al., 2008). The N-terminal regions of phytochromes are characterised by an unusual knot consisting of a region upstream of the PAS domain and a loop extending from the GAF domain (Wagner et al., 2005). The C-terminal region is a signalling domain, and cyanobacterial and bacterial phytochrome feature a HK domain. HK domains of phytochromes are expected to be similar to structurally well characterised Class I SKs based on the similar domain organisation and the sequence identity, for example the sequence identity between Agp1 HK and TM0853 (PDB code: 2C2A, Marina et al., 2005) is 30.32%. The C-terminal signalling domain of plant phytochromes has evolved to become a Ser/Thr kinase (Chen et al., 2004). Additionally, plant phytochromes feature two tandem PAS domains termed "Quail module" between the N-terminal PCM and the C-terminal signalling domain (Hughes, 2010).

The first structure determination of a phytochrome was reported in 2005 when the structure of the chromophore binding domain (CBD, Wagner et al., 2005) encompassing the PAS and GAF fragment of *D. radiodurans* phytochrome (DrBphP) was described. Since then many more phytochrome structures have been determined which are summarised in Table 1.2. Each phytochrome structures differ by variables such as the source organism, the type of the chromophore, the form in which its structure was determined, and the primary structure. A major breakthrough was made when the structures of the entire photosensory core modules (PCM) of phytochromes comprising the PAS, GAF, and PHY domains were simultaneously published in 2008. The structures of *Synechocystis* PCC6808 phytochrome 1 (Cph1) and *P. aeruginosa* phytochrome (PaBphP) were solved in their dark stable forms which are the Pfr- and the Pr-form, respectively (Essen et al., 2008, Yang et al., 2008). Cph1 and PaBphP crystallised as parallel and anti-parallel dimers, respectively. In the present thesis, the mode of dimerisation according to this nomenclature is described in Fig. 1.10. It is noted that the first report of crystallisation of a PCM fragment of a phytochrome was reported in 2006 by Scheerer et al., and a preliminary structure of *Agrobacterium*

tumefaciens phytochrome 1 was modelled based on the datasets described in the study from 2006 (Scheerer et al., 2010, Scheerer et al., 2006).

Despite the initial anticipation that elucidating the mechanism of photoactivation and signal transduction should be possible by comparing the structures of different phytochromes in the Pr and the Pfr-forms, it was soon realised that conformational differences cannot immediately be attributed to such differences due to numerous differences of variables between them. Cph1 is a classical phytochrome from a cyanobacterium with PCB as the chromophore, whereas PaBphP is a bathy-phytochrome from a γ-proteobacterium with BV as the chromophore. Conformational differences between Cph1 and PaBphP can be explained by any of the following variables, such as low identity of the primary structure (31.25%), genuine differences between Pr and Pfr-form, or structural features specific to the lineage of each phytochrome. The latter variable can be further divided into two types of classification - bacterial and cyanobacterial phytochromes, and the classical and the bathy-phytochromes.

Under natural conditions, there is a correlation between the type of covalently bound chromophore, the type of the phytochrome, and the position of the conserved cysteine residue (Section 1.3.2). It is possible however, to obtain an unnatural holoprotein with a covalently bound non-cognate bilin by introducing appropriate mutations. For example the Agp1 C20A/V249C mutant binds covalently to PCB, and crystallisation of such a holoprotein has been described (Scheerer et al., 2010). Structural information for this unnatural holoprotein may help to disentangle the specific effect the bilin imposes on the structure of the holoprotein.

At the start of this project, the source organisms from which the structures of phytochrome PCMs had been solved were limited to *Synechocystis* PCC 6803 and *P. aeruginosa*, therefore it has not been possible to disentangle the correlation between the variables and structural features. Subsequently more structural studies have been published, including the structures of *SyB*-Cph1 in both Pr- and Pfr-forms, and PaBphP in the mixed Pfr/Pr-forms. Both studies have some limitations, however. In case of the former, the *SyB*-Cph1 structures in solution were solved as the GAF domain fragment, and it is not ideal for deducing how the initially localised molecular changes are ultimately translated into the modulation of the kinase activity. For the latter, a logical question may be posed regarding crystal structures of a protein in mixed states. One

question could be that since some conformational change should be expected to occur during a photoconversion, how could then the simultaneous presence of two forms in a crystal have resulted in the refinement of a single entity? Should no conformational change be expected, then what is the mechanism of signal transduction from the N-terminal sensor domain to the C-terminal HK domain? Yang et al. (2009) interpreted the broad electron density map as the result of simultaneous presence of two forms, but working with such maps could lead to over-interpretation. While it is clear that each and every one of these studies have contributed to accumulation of our knowledge, it is still highly desirable for other phytochrome PCM structures to be solved, because the availability of multiple representatives for each type of phytochromes (classical and bathy) will enable the disentanglement of the variables contributing to conformational differences. Furthermore, structures of a single phytochrome in both Pr- and Pfr-forms are required in order to reveal the molecular trajectories during the photoconversion process. Structural information of a single phytochrome in both photochromic states will enable to carry out structural comparison without bias, and thus help to clarify the effect of photoconversion which culminates in signal transduction to the C-terminal histidine kinase.

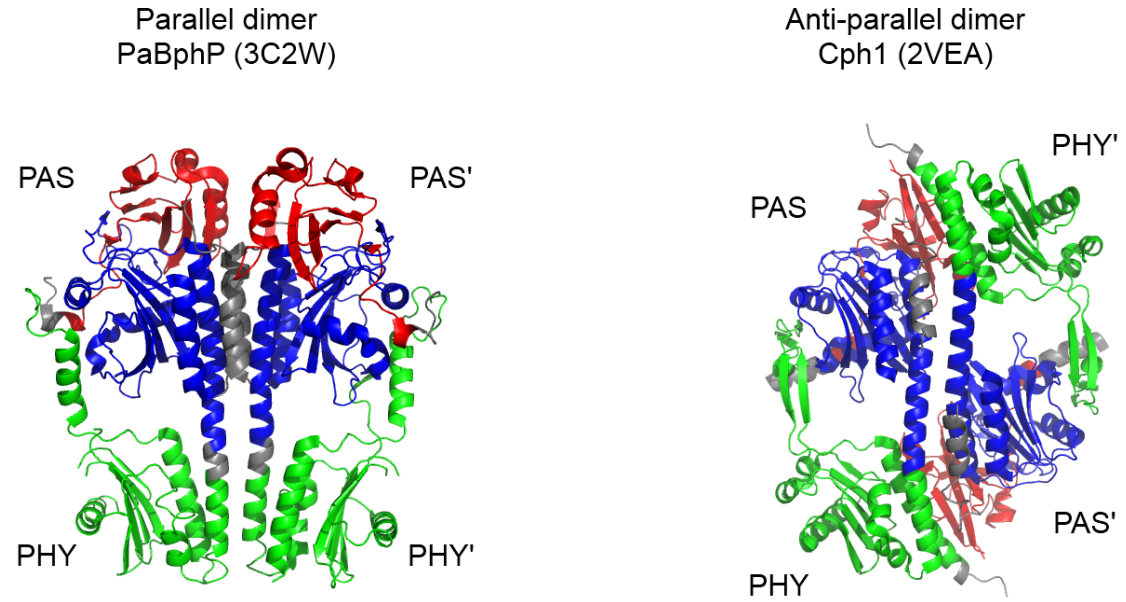


Figure 1.10 Two modes of dimerisation in crystal structures of phytochrome PCMs. PCM structures of Cph1 (PDB code: 2VEA) and PaBphP (PDB code: 3C2W) in ribbon presentation. PAS, GAF, and PHY domains as defined by the SMART server are in red, blue, and green, respectively. In the present thesis, a parallel dimer is defined by the corresponding domains at the longitudinal termini of two monomers in a dimer being *in cis* orientation. An anti-parallel dimer is defined by the corresponding domains at the longitudinal termini of two monomers in a dimer being *in trans* orientation.

Phytochrome	**Source organism**	**Included domains**	**The form**	**Chromophore**	**PDB code**	**Reference**
DrBphP	*D. radiodurans*	PAS, GAF	Pr	BV	1ZTU	Wagner et al. (2005)
DrBphP Y307S	*D. radiodurans*	PAS, GAF	Pr	BV	2O9C, 2O9B	Wagner et al. (2007)
RpBphP3	*R. palustris* CGA009	PAS, GAF	Pr	BV	2OOL	Yang et al. (2007)
Cph1	*Synechocystis* PCC 6803	PAS, GAF, PHY	Pr	PCB	2VEA	Essen et al. (2008)
PaBphP	*P. aeruginosa*	PAS, GAF, PHY	Pfr	BV	3C2W	Yang et al. (2008)
*Sy*B-Cph1	*Synechococcus* OS-B'	GAF	Pr	PCB	2K2N	Cornilescu et al. (2008)
PaBphP Q188L	*P. aeruginosa*	PAS, GAF, PHY	Pfr/Pr mixed	BV	3G6O, 3IBR	Yang et al. (2009)
*Sy*B-Cph1	*Synechococcus* OS-B'	GAF	Pr and Pfr	PCB	2KOI, 2KLI (2LB9, 2LB5)	Ulijasz et al. (2010)
PaBphP	*P. aeruginosa*	PAS, GAF, PHY	Pfr/Pr mixed	BV	3NHP, 3NHQ, 3NHT, 3NHU	Yang et al. (2011)
Cph1 Y263F	*Synechocystis* PCC 6803	PAS, GAF, PHY	Pr	PCB	3ZQ5	Mailliet et al. (2011)
DrBphP D207H and D207H/Y263F	*D. radiodurans*	PAS, GAF	Pr	BV	3S7N, 3S7O, 3S7P, 3S7Q	Auldridge et al. (2012)
RpBphP2	*R. Palustris* CGA009	PAS, GAF	Pr	BV	4E04	Bellini and Papiz. (2012)
RpBphP1	*R. Palustris* CGA009	PAS, GAF, PHY, PAS/PAC	Pfr	BV	4EHO	Bellini and Papiz. (2012)

Table 1.2 The list of phytochromes structurally determined at near-atomic resolution to the present in chronological order. PDB codes in parentheses indicate the accession code of the refined coordinates where available.

Response regulator	Source organism	PDB code	Reference
RcpA and RcpB	*Calothrix* PCC7601	1K68 and 1K66	Benda et al. (2004)
Rcp1	*Synechocystis* PCC 6803	1I3C and 1JLK.	Im et al. (2009)

Table 1.3 The list of structurally analysed cognate response regulators of phytochromes to the present in chronological order.

1.3.5 Signalling mechanism of phytochrome

Signal transduction mechanisms of SK in general were discussed in Section 1.2.4, but here phytochrome will be treated specifically.

Similar to SKs in general, our current understanding of signal transduction in phytochromes is rudimentary. This is partially due to the fact that only a few of the phytochromes have been characterised for their kinase activity in three different forms - as apoprotein, the Pr-form, and the Pfr-form. Such analyses have been carried out on Agp1, Agp2, Cph1, and PaBphP, and the consensus seems to be that the relative kinase activity is the highest for the apoprotein. As for the holoprotein, the dark adapted form has the higher kinase activity (Inomata et al., 2005, Karniol and Vierstra, 2003, Lamparter et al., 2001, Njimona and Lamparter, 2011, Psakis et al., 2011, Tasler et al., 2005). Originally PaBphP was reported to exhibit higher kinase activity in the Pfr-form (the dark adapted form), but the difference in kinase activities in the Pr- and the Pfr-forms were marginal (Tasler et al., 2005). The work by Tasler and co-workers is regarded as controversial, because an independent investigation using a mutant of PaBphP that cannot form the Pr-form concluded the Pr-form to have a higher kinase activity (Yang et al., 2011). At least one conclusion which can be drawn from the currently available observations is that the relative kinase activity, with respect to the bathochromic state, is different among phytochromes.

The modulation of kinase activity of a chimeric SK, created by fusing Cph1 and EnvZ (Levskaya et al., 2005), in response to light is seemingly paradoxical. The parental SK, Cph1 has higher kinase active in the Pr-form. The chimera is also more active in the Pr-form, but importantly the chimera contains a HAMP domain between the sensor and the HK domains. From its structure, the HAMP domain was implied to inverse the signal by reversing the direction of rotation of the linker helices (Hulko et al., 2006). This may be rationalised by the rotation model of signal transduction (Section 1.2.4), because according to the model the number of residues at the linker plays a key role in determining the kinase activity in response to the stimuli.

Whether the signal transduction mechanism is conserved between the plant phytochromes and *bona fide* SK-type (cyanobacterial and bacterial) phytochromes is unknown. The signal transduction in plant phytochromes may not be confined to the SK-type mechanism as some activity modes of plant phytochromes are not found in SK-type phytochromes, for example the reversible interaction between the Quail module of

Arabidopsis PhyB and PIF3 (Ni et al., 1999).

It is established for SK that the formation of a parallel dimer is required for its kinase functionality. As of now no full-length structure of a phytochrome with canonical domain organisation has been determined. The full-length structure of RpBphP1 has been determined (Bellini and Papiz, 2012b), but its off-set anti-parallel dimer mode is unlikely for canonical phytochromes for two reasons. Firstly, RpBphP1 is an atypical bacteriophytochrome which does not have a HK domain at its C-terminus. Experimental evidences support the canonical phytochromes to form a parallel dimer (Evans et al., 2006, Li et al., 2010). Secondly, no plausible mechanism can be proposed for HK activity for an anti-parallel phytochrome dimer, given that the formation of a dimer at the HK domain is a prerequisite for a kinase activity even for the ones that undergo *cis* autophosphorylation (Casino et al., 2009).

Two modes of inter-domain orientation have been proposed for the PCM and the HK domains in phytochromes. The HK domains of each monomer may be orientated either in *cis* or *trans* relative to the PCM dimer (Fig. 1.11). Indirect support for the *cis* orientation comes from a crystal structure of TM0853 including a HK with a PAS domain at its N-terminus (Fig. 1.5, Yamada et al., 2009). As described, the PAS domain shares a common fold with the GAF and PHY domains, therefore an analogy may be drawn between phytochromes and TM0853 regarding their domain orientations. Low resolution structures obtained by cryoelectron microscopy on DrBphP (Li et al., 2010) and small-angle X-ray scattering (SAXS) on RpBphP4 (Evans et al., 2006) do not rule out the possibility of inter-domain orientations in *trans*.

With regards to the orientation of the Quail module (two PAS domains) relative to the PCM, the near-full length crystal structure of RpBphP1 may be regarded as a proxy. The crystal structure includes the PCM followed by a PAS/PAC domain but does not include the C-terminal HOS domain (Bellini and Papiz, 2012b).

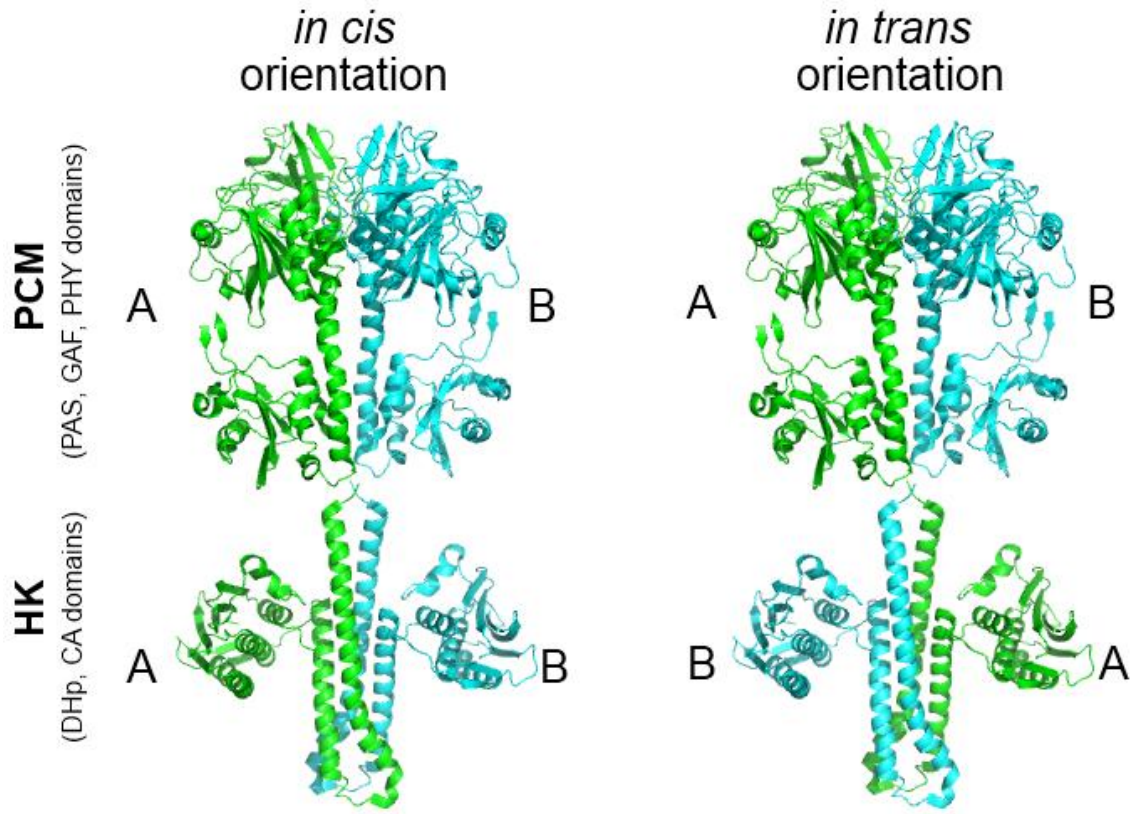


Figure 1.11 Two hypothetical modes of inter-domain orientations for the PCM and the HK domains in a full-length phytochrome. Structures are shown as cartoons. The preliminary structure of the WT M15-BV (Scheerer et al., 2010) is used as the N-terminal PCM and the C-terminal HK is from TM0853 (2C2A). A hypothetical full-length model of PaBphP with *cis* orientation has been presented in Yang et al. (2008). In Li et al. (2010) the *cis* and the *trans* orientations are referred to as mode I and II, respectively.

1.3.6 Applications of phytochromes in the future

Numerous applications of phytochromes have been suggested and demonstrated to date. Firstly, understanding the mechanistic details of phytochrome will lead to a better mechanistic understanding of the SK superfamily of TCS in general (Hughes, 2010). A system that can be reversibly switched on and off represents an improved model system compared to other chemical receptor SKs whose HK activity are modulated by diffusion-dependent stimuli.

Phytochromes are already known to control many physiological roles in plant, for example draught resistance (Gonzalez et al., 2012). It is anticipated that further insights into the molecular mechanism of phytochromes will enable more rational approaches to be pursued by genetic engineering to improve food and biofuel production (Ulijasz et al., 2010).

Structural studies in addition to biochemical characterisation of phytochrome mutants revealed the residues essential for the photoconversion. Fluorescent phytochromes with chromophores that do not permit photoconversion have been documented, for example an adduct with Phycoerythrobilin (PEB) or adducts with locked chromophores (Zienicke et al., 2011). Alternatively, experiments using holoproteins which has bound natural chromophore, but contains mutations at key residues, have also been shown to result in fluorescent phytochromes, possibly due to the loss of conformational freedom of the bilin chromophore around the D-ring (Samma et al., 2010). Fluorescent phytochromes display a maximum in the fluorescence spectrum towards the near-infra red (IR) region, therefore they are an attractive fluorophore for *in vivo* imaging because BV, the required cofactor, is available in the cell and light in the near-IR wavelength penetrates tissues more efficiently than light in the visible region (Lecoq and Schnitzer, 2011). Characterisation and expression of mutated phytochrome-based fluorophore in mammals have already been reported (Filonov et al., 2011, Shu et al., 2009).

Finally, the ability of phytochrome to translate the light signal into a biochemical signal has attracted its use as a component in the growing field of synthetic biology and optogenetics. The modular nature of SKs has enabled the creation of a chimeric SK from fragments of Cph1 and EnvZ as described in Section 1.2.4, which enabled the construction of a biological photographic system (Levskaya et al., 2005). In another example the ability of *Arabidopsis thaliana* PhyB is utilised to reversibly recruit phytochrome interacting factor 6 (PIF6) to spatially and temporally control a protein-protein interaction by light (Levskaya et al., 2009). Unique features of phytochrome will help to create diverse synthetic biological circuitries in the future along with a growing catalogue of biological parts (Registry of Standard Biological Parts, http://partsregistry.org/). A light-switchable promoter system has also been described (Shimizu-Sato et al., 2002). An increase in the overall structural information will undoubtedly help progress within the field of optogenetics (Moglich and Moffat, 2010).

1.3.7 Phytochromes in *Agrobacterium tumefaciens*

Agrobacterium tumefaciens is a non-photosynthetic plant pathogen, the causative agent of crown gall tumour in plants. Its use as means to transform plants has been established in the field of plant biotechnology (Christou, 1996), and its importance lead to the genome of C58 strain being sequenced, based on which two papers were published simultaneously(Goodner et al., 2001, Wood et al., 2001). Discovery of phytochrome

genes in a genome of a non-photosynthetic bacterium was a surprise, because initially no light was thought to be present in the natural sub-terrestrial habitat of *A. tumefaciens*. Later sunlight was found to be conducted by plant stem and roots (Sun et al., 2003) and it is also possible for *A. tumefaciens* to be exposed to the sun when on the surface of the soil (Scheerer et al., 2010). Discovery of phytochrome homologues in bacteria greatly accelerated biochemical characterisations of phytochromes as these are more amenable to overproduction with *E. coli* (Rockwell et al., 2006).

Phytochrome homologues from *A. tumefaciens* have been utilized as model phytochromes since their discovery, as in the present study. In this thesis a phytochrome termed Agp1 (alternatively referred to as AtBphP1, Karniol and Vierstra, 2003) is used as a model phytochrome, due to it being suitable for several reasons. It possesses a canonical domain organisation as well as being a classical phytochrome with standard photochemistry with the Pr-form being the dark adapted form (Lamparter et al., 2003). Additionally, the presence of another phytochrome Agp2 in the *A. tumefaciens* genome makes the system an interesting target for characterising the physiological roles *in vivo*, because Agp1 and Agp2 have antagonistic spectral properties and Agp2 possesses an atypical HK domain termed HWE-HK (Karniol and Vierstra, 2003, Karniol and Vierstra, 2004). Furthermore, as a HWE histidine kinase, Agp2 has with an RR domain at the C-terminus, which is in contrast to Agp1 with its typical domain organisation of a soluble SK.

Physiological roles of Agp1 and Agp2 in *A. tumefaciens* have been studied. Plant infectivity of *A. tumefaciens* was not altered by single or double knockout mutations of Agp1 and Agp2, regardless of the light condition. The growth rate of *A. tumefaciens* is reduced in both light and dark conditions when both phytochrome genes were knocked out (Oberpichler et al., 2006). In the following study, *A. tumefaciens* and other *Rhizobiaceae* species were found to exhibit the scotophobic (fear of darkness) response. Increased motility in darkness was found to originate from upregulation of flagella component proteins (FlaA and FlaB), but single or double knockout mutants of Agp1 and Agp2, as well as a knockout mutant of PhrA, a photolyase, exhibited identical swimming behaviour as the wild type, suggesting either the presence of an additional photoreceptor in the *A. tumefaciens* genome or phytochromes and photolyase to compensate for the absence of each other (Oberpichler et al., 2008). The future answer to this question may be related to the presence of PhrB, another homologue of photolyase which has been identified in *A. tumefaciens* genome (Oberpichler et al.,

2011). Recently, a study suggested the physiological role of Agp1 to be a temperature sensor as well as a light sensor (Njimona and Lamparter, 2011). The kinase activity of Agp1 decreased in response to increasing temperature while in general maintaining the relative order of kinase activity in decreasing order of apo-form, Pr-form, and Pfr-form. This reduction in kinase activity was not attributed to the denaturation of the protein as judged by the integrity of the absorption spectra obtained for the investigated range of temperatures. Furthermore, the temperature dependent formation of a novel Prx-form with a unique absorption spectrum was observed.

1.3.7.1 Start codon ambiguity

When interpreting the structural and biochemical data using the recombinant Agp1, it must be noted that there is an ambiguity regarding the location of the exact start codon. Subcloning of the Agp1 gene into an expression vector was first documented by Lamparter (2002) and the pair of primers used for amplification from the *A. tumefaciens* C58 genome is shown in Fig. 1.12.

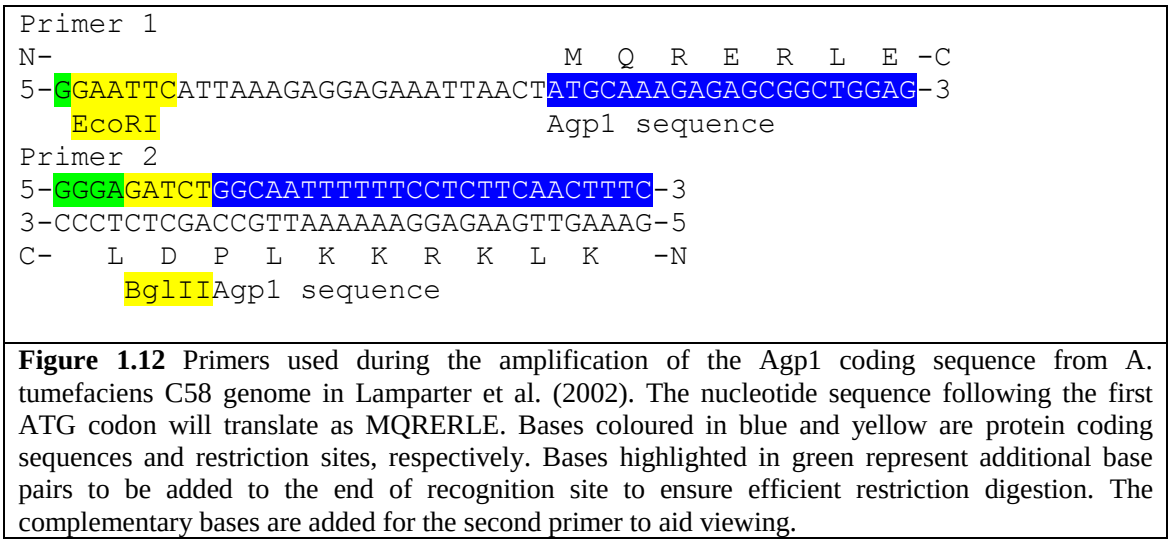
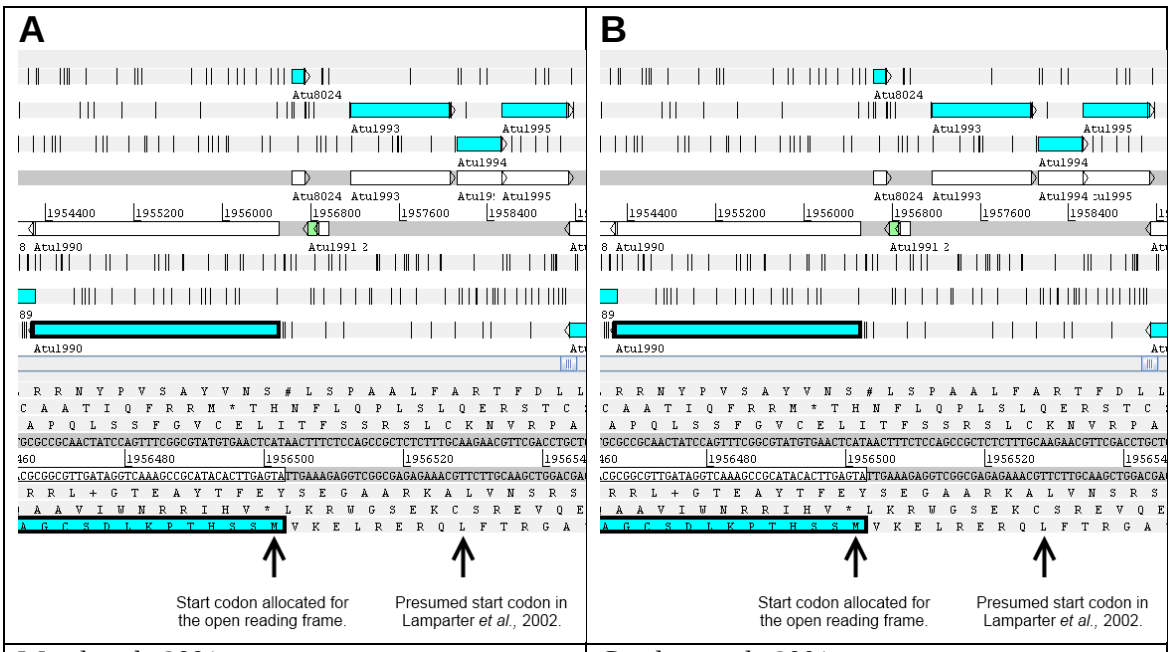


Figure 1.12 Primers used during the amplification of the Agp1 coding sequence from A. tumefaciens C58 genome in Lamparter et al. (2002). The nucleotide sequence following the first ATG codon will translate as MQRERLE. Bases coloured in blue and yellow are protein coding sequences and restriction sites, respectively. Bases highlighted in green represent additional base pairs to be added to the end of recognition site to ensure efficient restriction digestion. The complementary bases are added for the second primer to aid viewing.

Upon checking the open reading frame Atu1990, which corresponds to Agp1 in the circular chromosome of *A. tumefaciens* C58 from two independent genome projects published simultaneously (Goodner et al., 2001, Wood et al., 2001), it is clear that the position of the presumed start codon in Lamparter et al. (2002) is actually occupied by the codon TTG which is normally translated into leucine (Fig. 1.13). The annotated protein coding sequence starts from another downstream methionine which corresponds to Met10 according to the numbering by Lamparter et al. (2002) and as far as is understood, all subsequent studies using Agp1.



Wood et al., 2001 GenBank accession code: AE007869.2	Goodner et al., 2001 GenBank accession code: NC_003062

Figure 1.13 Sequences of the circular chromosome of A. tumefaciens C58 viewed with Artemis 13.2.1 (Rutherford et al., 2000). Sequences from two independent genome projects indicate the presence of a codon TGG in the position of the presumed start codon in Lamparter et al. (2002).

It is not unknown for a start codon to be TTG. In fact, about 3% of the start codons are TTG in case of *E. coli* K-12 (Blattner, 1997). No publication is yet available to justify the assignment of the TTG codon as the start codon of Agp1. Judging the exact location of the start codon from the sequence alone is more difficult than the stop codons.

Biochemical verification proved inconclusive. The absence of the first nine amino acids was shown not to drastically affect the overall properties of an Agp1 fragment termed M15 encompassing the PCM (Noack et al., 2007), which makes it difficult to judge whether the first nine amino acids according to the nomenclature of Lamparter et al., 2002 is a legitimate part of the endogenous protein. Crystallisation of a BV adduct of the M15 fragment of Agp1 yielded a slightly better result with the protein lacking the first nine residues (P. Scheerer, personal communication).

The most decisive way to clarify the issue will be to analyse the endogenously expressed Agp1 protein from *A. tumefaciens* C58 with Edman degradation or mass-spectrometry. Such work may be technically difficult owing to the low level of Agp1 being expressed in *A. tumefaciens* C58, which is estimated to be about ten molecules per cell (Oberpichler et al., 2006). Taking all the aforementioned factors into account, clarification of this issue does not justify the required resources in the present study. The majority of the crystallisation and biochemical analyses were carried out with the

constructs lacking the first nine amino acid residues unless otherwise stated, and they are identified by the suffix of "s" after the name of the constructs (e.g. M15s). For the sake of consistency, the residue numbering in the present study will correspond to the nomenclature originating from Lamparter et al., 2002 (i.e. the chromophore-binding Cys residue in Agp1 is at the 20^{th} position).

1.4 Aim of the project and thesis outline

The aim of this project was to obtain functional insights into phytochromes using structural information obtained via X-ray crystallography. Herein Agp1 was utilised as the model phytochrome. Agp1 was deemed a suitable target for crystallographic studies for several reasons. Firstly, Agp1 along with Agp2 have been subjects of intense biochemical and crystallographic analyses, and much data have already been accumulated for these proteins. Crystallisation of Agp1 PCM has been reported (Scheerer et al., 2006), and it is anticipated that the present study will be able to build upon the previous efforts. The new crystallographic structure obtained in the present study will be better rationalised and reconciled with data that have already accumulated for Agp1. Secondly, Agp1 is a classical bacteriophytochrome with a photocycle common to majority of phytochrome, hence its structural data is of wide relevance to other phytochromes. Moreover, at the start of the present study, Agp1 would have been the first classical bacteriophytochrome to be structurally determined as a PCM fragment, to the best of the author's knowledge. This is significant as the Agp1 structure has added substantial information to the combination of variables as found in the already known structures of phytochrome PCM. At the start of the present study, only two phytochrome PCM structures were available (2VEA, Essen et al., 2008; 3C2W, Yang et al., 2008). Two phytochrome structures were different in many aspects, but such structural difference could not simply be rationalised as the difference in the Pr and the Pfr-forms, because many other variables that contribute to the structure of those proteins were different too. The variables are for example the type of the chromophore (BV or PCB) and the lineage (cyanobacterial or bacterial). Hence it posed difficulty over assigning the specific effect each variable had on the structures of two different phytochromes. Structure of Agp1, a bacteriophytochrome with biliverdin in the Pr-form, will help to disentangle how each variable affects the structure of phytochromes.

To date, no phytochrome has had its structures of both Pr and Pfr forms solved. Structural comparison of two phytochrome PCM structures available at the start of the present study (Cph1 and PaBphP) reveals many differences between them, but ultimately it is not possible to deduce if the differences reflect the structural changes during a photoconversion in one phytochrome or if the differences are simply due to the low sequence identity. Whether the structure of Agp1 will turn out to resemble Cph1, PaBphP, or look like a mixture of both is impossible to predict from the sequence alone

as the sequence identities between all of them are similar (Table 1.4).

	Agp1	Cph1	PaBphP
Agp1		35.43%	33.33%
Cph1			31.25%

Table 1.4 Amino acid sequence identities between three phytochrome PCMs. Since the sequence identities are similar, it is difficult to predict whether the Agp1 structure will more closely resemble the Cph1 or the PaBphP structure. Sequences from the first residues (Met 1) to the last residues of the PHY domain as identified by the SMART server were aligned, and the residue numbers were as follows; Agp1 (1-504), Cph1 (1-513), PaBphP (1-498). Sequence identity was calculated with the number of aligned residues excluding deletions in the denominator.

Solving the structures of Agp1 in both Pr- and Pfr-forms will enable an unbiased structural comparison to be made in order to deduce the molecular trajectories during the photoconversion. Additionally, the molecular movement which induces the signal transduction, eventually culminating in histidine kinase activity modulation, might also be understood.

In the present study, structural biological studies have been associated with biochemical analyses to further characterise the difference between the Pr- and the Pfr-forms of Agp1. Recent publications now discuss the structures of phytochromes in detail especially in the context of photoconversion, but thus far apo-phytochromes have not been discussed in great detail. Analyses on apo-phytochrome fragments are also described in this thesis, because information regarding apo-phytochromes could describe how phytochromes acquire the ability to sense light. In this present study, Agp1 protein products of differing lengths were subjected to various biochemical analyses to pinpoint the location of conformational change as well as the quality of change. Additionally, since dimerisation is thought to be an essential prerequisite for a sensor kinase to carry out its kinase activity, the oligomerisation propensity of Agp1 was investigated.

This thesis is presented in a way such that after detailing the material and methods (Chapter 2), the structural analysis of Agp1 in the Pr-form is described in Chapter 3, followed by Chapter 4 that describes the attempts to structurally determine Agp1 in the Pfr-form. Next, the effects of the "Surface Entropy Reduction" strategy used during crystallisation of Agp1 in both Pr- and Pfr-forms are discussed in Chapter 5. Biochemical analyses to characterise the structural differences of Agp1 in different forms are presented in Chapter 6. Finally the overall summary is given in Chapter 7.

Chapter 2: Material and methods

2.1 Materials and instrumentation

2.1.1 Water

Water for most biochemical analyses was sourced from Purelab DV35 (ELGA), which supplies water of Type 2 quality of as defined by ASTM® (http://www.millipore.com/lab_water/clw4/tutorial&tabno=4). Water of further stringent quality for molecular biological applications was obtained from Purelab Ultra (ELGA), which was further treated with diethylpyrocarbonate **(**DEPC, Sigma) to minimise the effect of nucleases to the DNA samples.

2.1.2 Polymerase chain reaction

DNA was amplified by PCR using the C1000 thermal cycler from Biorad.

Phusion® high-fidelity DNA polymerase (Finnzymes) was used throughout the present study to amplify DNA, and was always used in combination with the HF buffer included in the kit. DMSO was used where indicated.

All primers were purchased from Eurofins.

2.1.3 DNA purification

The peqGOLD plasmid mini kit I C-line (Peqlab) was used to isolate plasmid DNA from overnight cultures of *E. coli*. The peqGOLD cycle-pure kit (Peqlab) was used to isolate plasmids or amplified DNA from PCR reactions.

2.1.4 DNA ligation

Blunt end DNA ligation was carried out with T4 DNA ligase (AppliChem).

2.1.5 Restriction endonucleases

All type-two restriction endonucleases used in the present study (NdeI, BamHI, and XhoI) were sourced from New England Biolabs.

2.1.6 Agarose gel electrophoresis

TAE buffer was used as a component of an agarose gel and the electrophoresis buffer.

10 × TAE buffer	Tris-HCl: 0.4 M Glacial acetic acid: 1.142% (v/v) EDTA: 0.01 M

2.1.7 Polyacrylamide gel electrophoresis

2.1.7.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

4 × Tris / SDS solution	Tris-HCl: 1.5 M (pH 8.8) SDS: 0.02% (w/v)
4.2% Stacking gel	30% Acrylamide / 0.8% Bisacrylamide: 14% (v/v) 4 × Tris / SDS solution: 25% (v/v)
4 × SDS loading solution	Tris-HCl: 200 mM (pH 6.8) DTT*: 400 mM SDS: 8% (w/v) Bromophenol blue: 0.4% (w/v) Glycerol: 40% (v/v)

*DTT was omitted from 4 × $SDS_{(non\text{-}reducing)}$ loading solution for non-reducing SDS-PAGE.

10 × SDS Running buffer	Tris-HCl: 0.25 M Glycine: 1.9 M SDS: 0.5% (w/v)

2.1.7.2 Native polyacrylamide gel electrophoresis

4 × Tris solution	Tris-HCl: 1.5 M (pH 8.8)
4.2% Stacking gel	30% Acrylamide / 0.8% Bisacrylamide: 14% (v/v) 4 × Tris solution: 25% (v/v)
Native-PAGE loading solution	Sucrose: 68% (w/v)
Detergent-free buffer	Tris-HCl: 0.05 M Glycine: 0.4 M EDTA: 0.001 M

2.1.7.3 Staining and de-staining of polyacrylamide gels

Coomassie staining solution*	Coomassie brilliant blue R250: 0.194 (w/v) Ethanol: 50% (v/v) Glacial acetic acid: 7% (v/v)

*Undissolved Coomassie brilliant blue was removed by Whatman No. 1 filter paper.

Destaining solution	Ethanol: 10% (v/v) Glacial acetic acid: 7% (v/v)

2.1.7.4 Molecular weight markers

Protein marker I (14-116, AppliChem) was used as the molecular weight standards in reducing and non-reducing SDS-PAGE. SN marker was also used, and it consisted of M15 (55.8 kDa) and M20 (34.7 kDa) apoproteins at the final concentration of 1 mg/ml. SN2 marker was identical to SN marker except for the constituting proteins were M15s (54.9 kDa) and M20s (33.5 kDa) apoproteins.

2.1.8 pH measurement

The pH of solutions was determined by a pH 210 microprocessor pH meter (Hanna). Alternatively, pH indicator strips (pH 0-14, VWR) were used to determine the pH of a small amount of solution.

2.1.9 Media for bacterial growth

All liquid media were sterilised by autoclaving prior to use. An antibiotic was added to the liquid media once it has cooled, prior to use.

LB and RB media

LB media	Tryptone: 1% (w/v) Yeast extract: 0.5% (w/v) NaCl: 0.5% (w/v)

For rich broth (RB), 0.2% (w/v) of D-glucose was added to the above.

SOC media

SOC media	Tryptone: 2% (w/v) Yeast extract: 0.5% (w/v) NaCl: 0.058% (w/v) KCl: 2.5 M Glucose: 0.4% (w/v)

Agar plate

Agar plates were used to obtain colonies of *E. coli* cells where an antibiotic may be added to select the cells bearing a plasmid that confers cells resistance, thereby allowing their growth.

Agar was added to LB media at a concentration of 1.5% (w/v) prior to sterilisation with an autoclave. Where required, an antibiotic was added to the media after the media has been autoclaved and cooled to 50-60 °C.

Antibiotics

All the plasmids used in the present study contain a gene encoding for β-lactamase, conferring host *E. coli* cell resistance to ampicillin, which was used as the selective antibiotic.

Ampicillin sodium salt was solubilised in 50% ethanol to the final concentration of 100 mg/ml. The working concentration of the ampicillin was 100 μg/ml.

2.1.10 Cell disruption

Collected *E. coli* cells were disrupted for protein extraction either by freeze/thaw cycles, sonication, or by sing a cell homogeniser, or a combination of those methods.

Sonication was carried out with a Vibra cell sonicator (Sonics). Cells were homogenised with an EmulsiFlex C3 cell homogeniser (Avestin).

2.1.11 Centrifugation

The following centrifuges were used to centrifuge samples at required centrifugal forces.

Desktop centrifuge: PerfectSpin 24 (Peqlab)

Medium volume centrifuge: 5840R (Eppendorf)

High speed/Large volume centrifuge: Avanti J-301 (Beckman Coulter)

Following rotors were used under appropriate situations in combination with the latter of the aforementioned centrifuges.

Large volume: JA-10 (Beckman Coulter)

High speed: JA-25.50 (Beckman Coulter)

2.1.12 Lyophilisation

Dehydration of samples was carried out with a Univapo 100H vacuum concentrator (Montreal Biotech Inc.).

2.1.13 Chromatography

Proteins were purified with the following fast protein liquid chromatography systems.

- ÄKTA FPLC with Frac-950 fraction collector, controlled by Unicorn 4.
- ÄKTA purifier with Frac-950 fraction collector, controlled by Unicorn 5.31.

All the buffers used during chromatography were filtered and degassed by a bottle top filter (Nalgene) and ME24 filter (Whatman) using a vacuum pump (Edwards).

Sep-Pak® Plus C18 cartridges (Waters) were used to buffer exchange biliverdin to remove HCl.

2.1.14 Protein concentration

Vivaspin 20 10,000 MWCO PES (Satorius) were used to concentrate protein samples.

2.1.15 Spectroscopy

2.1.15.1 UV-visible absorption spectroscopy

U-3010 UV-visible absorption spectrometer in combination with UV solutions 2.0 (Hitachi) was used to measure the absorption either as spectra or at a single wavelength.

A quartz cuvette (Hellma) was used for measuring absorption of DNA and protein samples, while polystyrene cuvettes (VWR) were used for measuring the optical density of *E. coli* culture.

2.1.15.2 Circular dichroism spectroscopy

CD spectroscopy was carried out with a Chirascan (Applied photophysics). Data was manipulated with Pro-Data viewer (Applied photophysics).

2.1.16 Dynamic light scattering

Dynamic light scattering (DLS) experiments were carried out with a Dyna Pro controlled by the UV solutions 2.0 software (Protein Solutions). A 45 µl sample cell was used to hold the sample during measurements.

2.1.17 Safe light

"Safe lights" were used to secure the visibility while preventing the photoconversion of phytochrome samples. Safe lights consist of either green filters or green LEDs, and are detailed in Appendix 1.

2.1.18 Protein crystallisation

Protein crystallisation screening plates were set up with a Mosquito robot (TTP labtech) controlled by the Mosquito software. Crystallisation screening kits were sourced from Molecular Dimensions, Hampton research, and Jena Biosciences. Either 2-drop 96-well

MRC plates or 3-drop polystyrene plates (Molecular Dimensions), both in sitting drop configuration, were used for screening crystallisation conditions. 24-well XRL plates (Molecular Dimensions) were used for optimisation of crystallisation conditions. All PEG precipitants were purchased from Molecular Dimensions. Crystallisation plates were incubated either in a temperature controlled room at 19 °C, or in either a temperature controlled incubator (Binder) or MD5-501 incubator (Molecular Dimensions). Crystallisation plates were inspected with a SM2-168 stereo microscope (Motic). Crystals were vitrified in liquid nitrogen using LithoLoops (Molecular Dimensions).

2.1.19 X-ray diffraction image collection

X-ray diffraction images were collected either with an in-house rotating anode X-ray generator or at the ESRF synchrotron, Grenoble.

The set-up for in-house experiments consisted of an RU-H3R RINT2000 series CE marking horizontal type rotor flex (RIGAKU, Japan), a Mar345 image plate scanner (Marresearch, Germany), and a cryostream cooler (Oxford cryosystems, UK). The rotating anode generator was operated at 50 kV and 100 mA.

For data collections at the ESRF synchrotron, the following beamline and detectors were used. ID14-4, Q210 CCD detector (ADSC); ID14-4 and ID23-1, Q315r CCD detectors (ADSC).

2.2 Methods

2.2.1 Water preparation

For molecular biological applications, the water obtained from Purelab Ultra (ELGA) was treated by DEPC. DEPC was added to the water at a concentration of 0.1% (v/v) and incubated for 3 hours at 37°C. The DEPC was inactivated by autoclaving the solution for 15 minutes at 121 °C.

2.2.2 Plasmid construct preparation

2.2.2.1 Polymerase chain reaction

PCR was used to synthesise the pNM19a-M20 construct for expressing the M20s gene (Agp1 10-313 and a hexahistidine tag) in *E. coli*. The entire template plasmid termed

pNM19a (Table 3.1) which bears the coding sequence of agp1 M15s (DNA 28-1512) was amplified, except for the sequence coding for the PHY domain to be omitted.

Lyophilised primers were solubilised in DEPC-treated water to a final concentration of 100 μM. Working solutions were made by further dilution with DEPC-treated water to the final concentration of 10 μM. All primer solutions were stored at -20 °C

Primers

The following primers were used for the generation of the pNM19a-M20 plasmid (Section 3.3.1).

Name	**Sequence**	**Tm (°C)**
pNM19a-M20F	5′CACCACCACCACCACCACTGAGATCCGGCTGC 3′(32 bp)	>75
pNM19a-M20R	5′CTGGGAACTCTGCTCCATCGCGATCCGCATCG 3′(34 bp)	74.6

PCR reaction mixture

The following composition was used for the generation of the pNM19a-M20 plasmid (Section 3.3.1).

Component	**Volume (μl)**
DEPC treated water	Adjusted for the reaction to be 50 μl
5 × HF buffer	10
dNTPs (10 mM)	1
Template parental plasmid	Adjusted to be 5 ng
Forward primer (10 μM)	1
Reverse primer (10 μM)	1
Phusion polymerase (2 U/μl)	0.5

Thermal cycles

The following thermal cycles were used to amplify the pNM19a-M20 plasmid (Section 3.3.1).

Stage	**Temperature (°C)**	**Time (seconds)**	**Cycles**
Initial denaturation	98	120	1
Denaturation	98	10	25
Annealing	65	15	
Extension	72	180	
Final extension	72	600	1
Storage	10	Hold	

2.2.2.2 PCR product purification

Amplified PCR product or plasmid DNA were treated with the peqGOLD cycle-pure kit (Peqlab) to remove oligonucleotides and the polymerase as well as to exchange the buffer. DNA samples in solution were subjected to the protocol specified by the kit. The DNA sample was always eluted from the column with DEPC-treated water, rather than the elution buffer, in order to make the sample more amenable to downstream applications.

2.2.2.3 Ethanol precipitation

Ethanol precipitation was used to purify plasmid DNA (Sambrook et al., 1989).

3.0 M of sodium acetate solution was added to the DNA sample in solution to the final concentration of 0.3 M and mixed. 2 volumes of ice-cold 100% ethanol were added and the solution was mixed. The solution was incubated at 4 °C for 30 minutes to allow for the DNA to precipitate. The sample was centrifuged at 12,000 × g for 10 minutes at 4 °C. The supernatant was removed, and the sample tube was filled with a half volume of 70% ice-cold ethanol. The sample was centrifuged again with the same parameters, and the supernatant was removed. The tube was allowed to dry at room temperature with the lid open until no trace of liquid was identified. The DNA pellet was dissolved with 50-100 μl of DEPC-treated water.

2.2.2.4 Miniprep

Plasmids were purified predominantly from *E. coli* XL1-blue strain cells, bearing the plasmid of interest. 1-5 ml of overnight culture was centrifuged at 10,000 × g for 1 minute at the room temperature. The collected cell pellet was subjected to the alkaline lysis protocol given in the peqGOLD plasmid mini kit (Peqlab). The plasmid DNA was always eluted from the column with DEPC-treated water, rather than the supplied elution buffer.

2.2.2.5 Restriction digestion

Restriction digestion of DNA was carried out under the presence of the appropriate NEB buffer, and the DNA sample was supplemented with bovine serum albumin (BSA) where necessary (BglI, NEB3; DpnI, NEB4; NheI, NEB2, +BSA; XhoI, NEB4, +BSA).

2.2.2.6 T4 Ligation

Blunt end ligation was carried out with T4 ligase. The reaction was carried out for 1:30 hours at 20 °C with the following composition.

Component	**Volume (µl)**
DNA sample	25
10 × Buffer	3
T4 ligase (1,000 units/ml)	2

2.2.2.7 Agarose gel electrophoresis

The length and the quantity of double-stranded DNA fragments were estimated by agarose gel electrophoresis. Agarose (Merck) was dissolved in 1 × TAE buffer to a concentration of 0.7% (w/v), and ethidium bromide solution (Fisher) was added to the final concentration of 50 µg/ml. The agarose solution was cast into a tank with a comb (Biorad) until set. DNA samples were supplemented with 6 × DNA loading buffer (Peqlab) and subjected to electrophoresis at a constant voltage of 2 V/cm until sufficient separation was achieved. Bands in gels were visualised under an illumination with UV light platform (Uvitec).

2.2.2.8 DNA quantification

The absorption at 260 nm allows the quantitation of DNA in the sample, where an OD of 1 equals to 50 ng/µl for double-stranded DNA (Sambrook et al., 1989). The Absorption at 280 nm was also measured to calculate the A260/A280 ratio used to quantify the sample purity. The typical ratio ranged from 1.7 to 2.0.

2.2.2.9 QuikChange™ mutagenesis

Site-directed mutagenesis was carried out according to the QuikChange™ mutagenesis protocol (Agilent Technologies) albeit individual enzymes were purchased separately from different manufactures (DNA polymerase: Finnzymes, DpnI endonuclease: New England Biolabs).

To facilitate the conformation of successful mutation, the following recognition sites for restriction enzymes were additionally incorporated. SER2, NheI; SER3, BglI; SER6, BglI (Table 2.1).

Name	Sequence	Tm (°C)	GC (%)
SER2 F	5′ccgcctggtattccggcgCTgCTagcaatttccgatatgcc 3′	79.28	58.54
SER2 R	5′ggcatatcggaaattgctAGcAGcgccggaataccaggcgg 3′	79.28	58.54
SER3 F	5′cgatatgcctgggccgCaGCgGCactcgatgtttcggcgc 3′	81.82	65.91
SER3 R	5′gcgccgaaacatcgagtGCcGCtGcggcccaggcatatcg 3′	81.82	65.91
SER6 F	5′ggccgccgccgCGGCatgggttgatggcctgc 3′	79.94	78.13
SER6 R	5′gcaggccatcaacccatGCCGcggcggcggcc 3′	79.94	78.13

Table 2.1 List of primers used during QuikChange mutagenesis to generate expression plasmids with SER mutations. Primer type is indicated after the name given to each mutant. F, forward primer; R, reverse primer. Region highlighted in yellow is where the mutations have been introduced. Underlined regions correspond to the recognition sites of the restriction enzymes introduced additionally to the designed mutations. SER2, NheI (G/CTAGC); SER3/6, BglI (GCCNNNN/NGGC).

Mutagenesis reaction was always performed at 50 µl per reaction. Composition of the reaction is as follows.

Component	Volume (µl)
DEPC treated water	Adjusted for the reaction to be 50 µl
5 × HF buffer	10
10 mM dNTPs (10 mM)	1
Template parental plasmid	Adjusted to be 50 ng
DMSO (100%)	1.5
Forward primer (10 µM)	1
Reverse primer (10 µM)	1
Phusion polymerase (2 U/µl)	0.5

The thermal cycle used for mutagenesis is as follows.

Stage	Temperature (°C)	Time (seconds)	Cycles
Initial denaturation	98	120	1
Denaturation	98	10	25
Annealing	55	60	
Extension	72	206	
Final extension	72	600	1
Storage	10	Hold	

Since site-directed mutagenesis was always carried out on variations of the pNM19a plasmid, the extension time is the same for all the mutants.

2.2.2.10 DNA sequencing

Despite a proof-reading capability of Phusion polymerase, it is possible for a mutation to be inadvertently introduced as a result of replication error. In order to verify the correctness of the sequence of gene of interest in the plasmid, or to confirm the success

of site-directed mutagenesis, dideoxy chain termination sequencing was carried out. Plasmid DNA samples were prepared according to the recommendation of Eurofins (15 μl of 50-100 ng/μl in double distilled water or Tris-HCl [pH 8.0 – 9.0]) before being shipped. Primers used for sequencing offered by Eurofins are designed to anneal to either the T7 promoter or the T7 terminator sequence of pET plasmids, and will be referred to as T7 forward and T7 reverse, respectively.

2.2.3 Protein production

2.2.3.1 Overnight culture

1-10 ml of LB media, containing appropriate antibiotic if necessary, was inoculated with either a single colony of *E. coli* cells, a glycerol stock or another liquid culture. The culture was incubated overnight at 30-37 °C while shaken at 200-220 rpm.

2.2.3.2 Chemically competent cell preparation

Various strains of *E. coli* cells were made chemically competent with divalent cations, rendering them amenable to be transformed. 100 ml of LB media containing appropriate selected antibiotic was inoculated with an overnight culture at a 1/200 dilution. Cells were allowed to grow at 37 °C with 200 rpm shaking until OD600 = 0.5. Cells were collected by centrifugation at 5,000 × g for 10 minutes at 4 °C. Cells were resuspended in 80 ml of ice-cold 100 mM $MgCl_2$ solution, and were centrifuged again. Pelleted cells were resuspended in 40 ml of ice-cold 50 mM $CaCl_2$ solution and incubated on ice for 30 minutes. Cells were centrifuged, and finally the cells were suspended in 4 ml of ice-cold solution containing 50 mM $CaCl_2$ and 15% (v/v) glycerol. Resuspended cells were aliquoted into 100 μl portions and flash frozen in liquid nitrogen. Frozen cells were stored at -80 °C.

2.2.3.3 Transformation

Chemically competent *E.coli* cells were transformed by heat-shock to introduce a plasmid into the cell. An aliquot of chemically competent cells was thawed on ice. 1 pg-100 ng of plasmid DNA in 1-5 μl volume was added to the cells and gently mixed with pipette tips on ice. The mixture was incubated on ice for 30 minutes. The mixture was heat-shocked at 42 °C for 90 seconds, and then cooled on ice for 2 minutes. 1 ml of SOC media without an antibiotic was added and the culture was incubated at 37 °C for 1 hour while shaken at 220 rpm. The culture was centrifuged at 3,000 × g for 3 minutes,

and 200 µl of the supernatant was pipetted and discarded. Additional 200 µl of fresh SOC media was added, and the cell pellet was resuspended, and plated onto LB agar plate containing the appropriate selective antibiotic.

2.2.3.4 Glycerol stock preparation

Glycerol stocks of *E. coli* cells were created to be maintained indefinitely. An overnight culture was supplemented with 20% (v/v) glycerol prior to flash freezing in liquid nitrogen. The glycerol stocks were stored at -80 °C.

2.2.3.5 Preparation of bilin chromophore

2.2.3.5.1 Biliverdin

Biliverdin hydrochloride is supplied in an acidic form that upon addition of biliverdin solution in organic solvent to protein solution causes the Agp1 to precipitate irreversibly, possibly due to the denaturation of the protein. Therefore, it was necessary to remove the acid by exchanging the buffer.

C18 column (Waters) was equilibrated with 5 ml of methanol. Roughly 5 mg of biliverdin was dissolved in 200 µl methanol, and was applied to the column. The column was washed with 30 ml of H_2O to remove HCl. The H_2O was then removed from the column by blowing air through it with a syringe. Biliverdin was eluted from the column with methanol. Eluted biliverdin in methanol was dried by a vacuum concentrator, and the dry pelleted biliverdin void of HCl was dissolved in DMSO, and stored at -80 °C.

2.2.3.5.2 15Ea

A synthetic bilin compound termed 15Ea which has the same stereochemistry between the C- and the D-rings as the natural bilin chromophore in the Pfr-form was assembled with the Agp1 apoprotein to create a holoprotein with similar biochemical characteristics to that of Agp1 in the Pfr-form (Inomata et al., 2005).

15Ea was supplied as dry powder. The sample was dissolved in DMSO and stored at -80 °C.

2.2.3.6 Production of recombinant proteins

2.2.3.6.1 Small scale test expression

Small scale test expressions were often carried out prior to large scale protein production. 1 – 10 ml of LB-Amp was inoculated with an overnight culture of the appropriate *E. coli* strain bearing the plasmid construct with 1/200 dilution. The cells were grown at 37 °C and 200 rpm until A_{600} reached 0.6. The culture was cooled at the target temperature, typically at 18 °C, for 1 hour while continued to be shaken. Protein expression was induced with 100 μM IPTG unless otherwise specified. The culture was incubated for 18 hours, and the cells were collected by centrifugation at 5,000 × g for 15 minutes. 1 g of collected cells was suspended in 10 ml of Basic buffer (300 mM NaCl, 50 mM Tris-HCl pH 7.8 at 20 °C, 5 mM EDTA). Cells were lysed either by 3 cycles of freeze-thaw and/or 5 cycles of sonication at 10 W for 30 seconds followed by a cooling period for 30 seconds. The cell lysate was clarified by centrifugation at 12,000 × g for 15 minutes. The supernatant was transferred into a fresh tube, and the insoluble pellet was resuspended in 500 μl of 4 ×SDS loading solution by vortexing. The resuspended pellet was heated at >80 °C for 5 minutes in a hot water bath. Soluble and insoluble fractions were analysed by SDS-PAGE.

2.2.3.6.2 Large scale growth and induction

Typically, 800 ml of LB media with appropriate antibiotic were prepared in 2L baffled flasks. The media was inoculated with an overnight culture at 1/200 dilution. The culture was incubated at 30-37 °C with 180-200 rpm until the OD600 reached 0.6. The cultures were cooled for 1 hour at target induction temperature of 16-18 °C while continued to be shaken. Expression of the gene of the target protein was induced with IPTG at the final concentration of typically 50 μM. The culture was incubated for further 18 hours.

E. coli cells from large scale cultures were collected by centrifugation at 6,000 × g for 20 minutes at 4 °C. Collected cells were typically stored at -80 °C.

2.2.3.7 Purification and assembly of recombinant proteins

2.2.3.7.1 Cell disruption and lysate clarification

Target proteins in the cytoplasm of *E. coli* cells were extracted by mechanical disruption methods. Pelleted cells were suspended in 60 ml Affinity buffer A (300mM NaCl, 50 mM Tris-HCl pH 7.8 at 20 °C, 10 mM Imidazole) and were disrupted by three passages though a cell homogeniser (twice at 5000 psi and once at 10,000 psi). The cell lysate was clarified by two rounds of centrifugation at 18,000 × g at 4 °C.

2.2.3.7.2 Affinity chromatography

As a first step of purifying the target protein, affinity chromatography was carried out. All isolated proteins in the present study have a C-terminal 6 histidine tag, and are purified by Ni^{2+} affinity chromatography.

Clear lysate was applied to a pre-equilibrated HiTrap IMAC HP 5ml column (GE Healthcare) loaded with Ni^{2+} using a P1 peristaltic pump (GE Healthcare). Weakly bound protein was removed by applying Affinity buffer A until the OD280 of the flow-through reached below 0.1 or the total volume of flow-through reached 10 CV.

Proteins strongly bound to the column were eluted by applying Affinity buffer B (300mM NaCl, 50 mM Tris-HCl pH 7.8 at 20 °C, 250 mM Imidazole) until the OD280 of the flow-through dropped below 0.1 or the total volume of flow-through reached 6 CV.

In order to remove imidazole from the protein eluate, the protein was precipitated by adding three parts of Ammonium sulphate buffer (3.3M $(NH_4)_2SO_4$, 20 mM Tris-HCl pH 7.8 at 20 °C, 5 mM EDTA) to two parts of eluted protein. The precipitated protein was collected by centrifugation at 6,000 × g for 15 minutes at 4 °C. Protein was solubilised with very small amounts of Assembly buffer (300 mM NaCl, 50 mM Tris-HCl pH 7.8 at 20 °C, 5 mM EDTA, 10% Glycerol). Any remaining insoluble matter was removed by centrifugation with the same parameters as mentioned above. The apoprotein was aliquoted in 0.5-1.0 ml portions, and flash frozen in liquid nitrogen and stored at -80 °C.

2.2.3.7.3 *In vitro* assembly of the Agp1 holoprotein

One of the defining features of *bona fide* phytochrome is autolyase activity to form a covalent bond with a bilin chromophore (Rockwell et al., 2006). Since Agp1 is a bacterial phytochrome that naturally uses biliverdin as the chromophore, it was possible to assemble the Agp1 holoprotein using commercially available biliverdin (Section 3). The Agp1 apoprotein was assembled with 15Ea to form the holoprotein in a Pfr-like form.

DTT was added to the Agp1 apoprotein to the final concentration of 10 mM and allowed to incubate on ice for 15 minutes. Protein was assembled with chromophores with molar ratios of 1:1.2 and 1:1 for BV and 15Ea, respectively. A bilin solution in DMSO was pipetted into protein solution, and mixed by pipetting up and down. *In vitro* assemblies and all subsequent downstream processes were carried out under the minimum exposure to the safelight. The assembly was allowed to proceed at 4 °C for 1 hour and 14 hours for BV and 15Ea adduct, respectively.

2.2.3.7.4 Ion exchange chromatography

Where the *in vitro* assembly of the Agp1 holoprotein was not efficient, apoprotein and holoprotein were separated by ion exchange chromatography (IEX). The buffer of a protein sample was exchanged to IEX buffer A (20 mM Tris-HCl pH 7.8 at 20 °C, 1 mM EDTA) by dialysis to ensure efficient binding of the protein to the stationary phase by lowering salt concentrations. Salt concentrations were further lowered by diluting the dialysed sample with IEX buffer A until to the volume of 10 ml.

Protein was bound to a pre-equilibrated HiTrap Q HP 5 ml column (GE Healthcare) using either a peristaltic pump or the sample injection loop of the FPLC. Protein was eluted by a linear gradient from IEX Buffer A to IEX buffer B (500 mM NaCl, 20 mM Tris-HCl pH 7.8 at 20 °C, 1 mM EDTA) over 15 CV and 30 CV for analytical and preparative purpose, respectively. Flow rate of the FPLC was set at 0.4-0.5 ml/min or 3 ml/min for preparative and analytical purpose, respectively. Eluted proteins were collected by a fraction collector.

For some chromatographic analyses, the quantities of protein samples were unified on the basis of absorption at 280 nm. The absorption of the sample was multiplied by its volume (ml) to derive a value with a unit of "mAU×ml", where the length of the cuvette

(*d*) is constant at 1 cm. This approach enables to unify the areas of the elution peaks in chromatograms.

2.2.3.7.5 Hydrophobic interaction chromatography

As an alternative way of separating holoprotein from apoprotein, hydrophobic interaction chromatography (HIC) was used. The concentration of ammonium sulphate in the protein sample was adjusted to 500 mM by adding the Ammonium sulphate buffer (Section 2.2.3.7.2). Protein was applied to a pre-equilibrated HiTrap Phenyl HP 5 ml column (GE Healthcare) either by a peristaltic pump or by using the sample injection loop of the FPLC. Protein was eluted with a linear gradient from HIC buffer A (500 mM $(NH_4)_2SO_4$, 300 mM NaCl, 20 mM Tris-HCl pH 7.8 at 20 °C, 1 mM EDTA) to HIC buffer B (20 mM Tris-HCl pH 7.8 at 20 °C, 1 mM EDTA) over 15 CV at 3 ml/min.

2.2.3.7.6 Size exclusion chromatography

Either HiPrep 16/60 Sephacryl S-200 High Resolution column or Superdex 200 10/300 GL column (GE Healthcare) were used for size exclusion chromatography (SEC). The former was predominantly used for the preparative purpose, and the latter was used for both preparative and analytical purposes. Protein samples were always subjected to SEC regardless of the preceding purification steps. The aims are to exchange the sample buffer, remove contaminations with different molecular weights, and to isolate a conformationally homogeneous species of the eluate.

The protein sample volume was adjusted to the volumes of the sample injection loop, which were either 100 µl or 2 ml. The sample was applied to the column with a sample injection loop and the protein was purified using an isocratic elution with a flow rate of 0.5 ml/min. Unless otherwise specified, the SEC buffer (20 mM Tris-HCl pH 7.8 at 20 °C, 1 mM EDTA) was used during SEC runs.

2.2.3.8 Concentration of protein

The membrane of a Vivaspin concentrator was pre-rinsed using H_2O to remove traces of glycerol and sodium azide. The concentrator was equilibrated with a suitable buffer prior to use. The maximum volume was 14 ml for the use with a centrifuge with a fixed angled rotor. The concentrator was centrifuged at 3,000-8,000 × g until the target concentration of the protein was achieved.

2.2.4 Biochemical analyses

2.2.4.1 Polyacrylamide gel electrophoresis

In polyacrylamide gel electrophoresis (PAGE), a negatively charged proteins migrate through a porous matrix of polymerised acrylamide and bisacrylamide gel, and the rate of migration is determined predominantly by the molecular weight of the sample. The gels were invariably stained by Coomassie brilliant blue to visualise the protein following the electrophoresis.

2.2.4.1.1 Reducing SDS-PAGE

Reducing SDS-PAGE was used to estimate the molecular weight of a target protein and the purity of the sample. Under the presence of saturating SDS, SDS binds to polypeptide chains at 1.4 g per gram of polypeptide. Uniform negative charge linearises the polypeptide chain to destroy the tertiary structure as well as the intermolecular aggregation. SDS binds almost proportionally to the length of the polypeptide regardless of the protein, so the resulting mass/charge ratio is almost always constant, thus making the length of polypeptide the dominant factor in determining the rate of migration through the gel. Furthermore, the reducing agent will break intra/inter-molecular disulphide bond(s).

Gel cast assembly was set up according to the manufacturer's instruction. In essence the protocol was in accordance with the standard procedure (Sambrook et al., 1989). The target concentrations of acrylamide/bisacrylamide in the resolving layer of gels were achieved according to Table 2.2. The samples were heated at 80 °C for 5 minutes in 1 × SDS loading solution. The electrophoresis was carried out at a constant voltage of 15 V/cm for 1 and a half hours under the presence of SDS running buffer.

	Final concentration of bis/acrylamide in separating gel (%)				
Stock solution	**8**	**9**	**10**	**11**	**12**
30% Acrylamide 0.8% Bisacrylamide	1.33	1.5	1.67	1.83	2
4 × Tris/SDS solution	1.25	1.25	1.25	1.25	1.25
H_2O	2.42	2.25	2.08	1.92	1.75

Table 2.2 Solutions and their volume for the resolving layer of SDS-PAGE gels with varying final concentration of bis/acrylamide. Volumes of the solutions are indicated in ml.

2.2.4.1.2 Non-reducing SDS-PAGE

Normally a reducing agent is included in the sample loading solution in SDS-PAGE, which breaks intra/inter-molecular disulphide bonds. By omitting the reducing agent from the loading solution, disulphide bonds are conserved and the rate of migration will be affected by disulphide bonds. SDS that linearises polypeptide chains still ensure that the rate of migration is mostly determined by the molecular weight of the polypeptide.

The protocol was identical to that for SDS-PAGE (Section 2.2.7.1.1), except that the reducing agent was omitted from the sample loading solution (Section 2.1.7.1) and the sample was not heated.

2.2.4.1.3 Native-PAGE

A native polyacrylamide gel electrophoresis may be carried out without the presence of neither SDS nor the reducing agent, thus the tertiary/quaternary structures of the protein is preserved during the separation. The resulting separation may be difficult to interpret since the rate of migration depends on the surface charge of the protein under the pH of the buffer when the separation is carried out, which is inconsistent among different proteins. Preservation of quaternary structure allows one to observe the oligomerisation of proteins under a condition closer to that in the *in vivo* environment. Therefore a native-PAGE may be used to study the conformational homogeneity of the protein.

A native-PAGE gel was casted with the same protocol as for an SDS-PAGE gel except for the lack of SDS and the presence of glycerol in the resolving layer (Table 2.3). The stacking layer of a native-PAGE gel was more alkaline than that for an SDS-PAGE gel (Section 2.1.7.2). The sample and native-PAGE loading solution were mixed at a volume ratio of 2:1. The electrophoresis was carried out at a constant voltage of 15 V/cm for 1:30 hours under the presence of Detergent-free buffer.

	Final concentration of bis/acrylamide in separating gel (%)				
Stock solution	**8**	**9**	**10**	**11**	**12**
30% Acrylamide 0.8% Bisacrylamide	1.33	1.5	1.67	1.83	2
4 × Tris solution	1.25	1.25	1.25	1.25	1.25
100% Glycerol	0.5	0.5	0.5	0.5	0.5
H_2O	1.92	1.75	1.58	1.42	1.25

Table 2.3 Solutions and their volume for the resolving layer of native-PAGE gels with varying final concentration of bis/acrylamide. Volumes of the solutions are indicated in ml.

2.2.4.2 UV-visible absorption spectroscopy

Absorption of translucent materials was quantified with UV-visible absorption spectroscopy, from which the molar concentration could subsequently be calculated with the extinction coefficient and the Beer-Lambert's law equation.

$$A = \varepsilon \times c \times l$$

Where A is the absorbance in absorption unit, ε is the extinction coefficient in $M^{-1}cm^{-1}$, *c* is the concentration in M, and *l* is the path length in cm.

2.2.4.3 Calculation of bilin chromophore concentration

The absorption at the following wavelengths and coefficients were used for calculating bilin concentration in HCl (5%)/methanol.

Compound	**Wavelength (nm)**	**ε ($mM^{-1}cm^{-1}$)**	**Reference**
BV	696	30.80	Lamparter et al., 2001
15Ea	679	51.30	Zienicke et al., 2011

2.2.4.4 Calculation of protein concentration

Extinction coefficients shown in Table 2.4 were used for calculating protein concentration. For the apoprotein without any disulphide bond, the calculated extinction coefficient without any disulphide bond from the Protein calculator server (http://www.scripps.edu/~cdputnam/protcalc.html) was used. The extinction coefficient of a holoprotein was derived by adding the ε280 of the appropriate bilin chromophore to the ε280 of the apoprotein. Absorption spectra of BV and 15Ea chromophores were taken in order to determine their ε280 (Fig. 2.1).

The extinction coefficient at the reference wavelength was scaled according to the absorption measurement at 280 nm and the reference wavelength.

The absorption value at 280 nm and that of the wavelength from which the reference extinction coefficients have been quoted were used to calculate the extinction coefficient for both BV and 15Ea.

In the present study the effect of covalent bond formation between protein and chromophore to A280 was neglected.

Protein / compound	Protein description	ε280 nm ($mM^{-1}cm^{-1}$)
Full-length	1-745 + ArgSer (BglII site) + hexahistidine tag	98290
M15	1-504 + hexahistidine tag	77950
M15s	10-504 + LeuGlu (XhoI site) + hexahistidine tag	77950
M15 C20A	1-504 C20A + LeuGlu (XhoI site) + hexahistidine tag	77950
M20	1-313 + hexahistidine tag	34280
M20s	10-313 + hexahistidine tag	34280
SER2	M15s E86A/E87A	77950
SER3	M15s E96A/K97A/K98A	77950
SER6	M15s E336A/K337A	77950
SER26	M15s E86A/E87A/E336A/K337A	77950
SER36	M15s E96A/K97A/K98A/E336A/K337A	77950
SER23	M15s E86A/E87A/E96A/K97A/K98A	77950
SER236	M15s E86A/E87A/E96A/K97A/K98A/E336A/K337A	77950
BV		20.963
15Ea		11.013

Table 2.4 Extinction coefficients (ε) at 280 nm of proteins and compounds used in the present study.

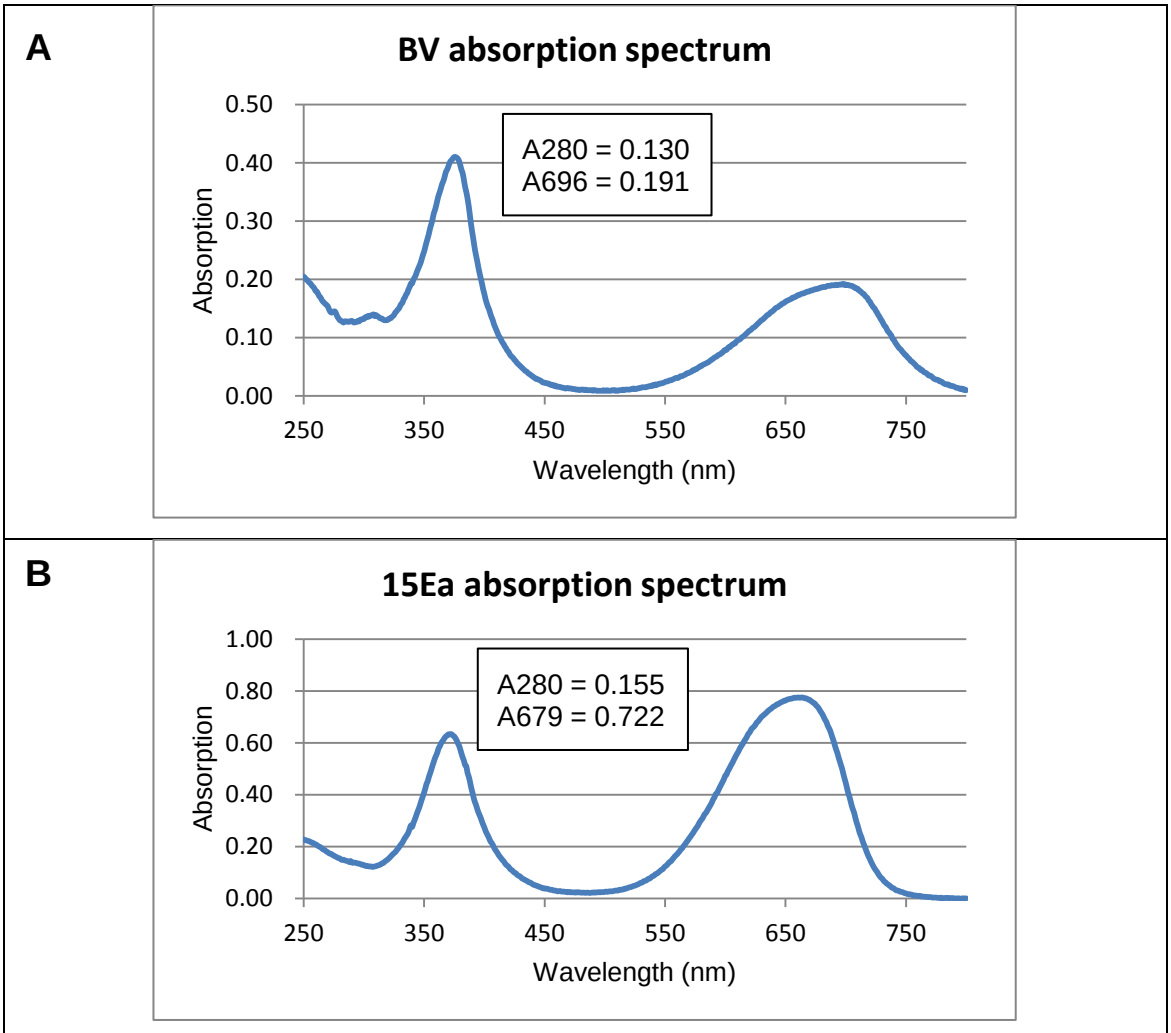


Figure 2.1 Absorption spectra of BV and 15Ea. Absorption spectra were obtained by diluting the bilin solution in DMSO to an arbitrary concentration with HCl (5%)/MeOH. A_{280}/A_{Qmax} ratios were obtained from spectra and were used to calculate extinction coefficients at 280 nm. A_{280}/A_{Qmax} for BV and 15Ea were 0.681 and 0.215, respectively. The extinction coefficients at 280 nm for BV and 15Ea were 20.963 and 11.013 mM^{-1}cm-1, respectively.

2.2.4.5 Circular dichroism spectroscopy

The secondary structure content of a protein was probed with CD spectroscopy, based on which comparative studies may be carried out to deduce photochromic state (Pr and Pfr) dependent conformational differences.

The buffer of protein samples were exchanged to CD buffer (20 mM Na phosphate pH 7.8, 0.5 mM EDTA) either by dialysis or with Vivaspin protein concentrators (Satorius). Protein concentration was adjusted to ≈0.1 mg/ml for the CD measurement and the concentration was quantified prior to the CD measurement in order to enable the calculation of molar circular dichroism after the measurement. The cuvette with a path length of 0.1 mm holding the protein sample was constantly orientated in the same direction in the sample chamber of CD spectrometer in order to minimize the variation being introduced during an experiment. Three independent measurements were carried out where the delta absorbance was measured between 190-260 nm. The spectra were subsequently averaged and the averaged spectrum of the CD buffer was subtracted. The spectra were adjusted for the reading at 260 nm to be 0 mdeg, and finally the spectra were smoothed with window value of 10 nm.

The spectra in mdeg as a function of wavelength were converted to mean residue ellipticity with the following equation.

$$\Delta\varepsilon = \theta \times \frac{0.1 \times MRW}{P \times CONC \times 3298}$$

where MRW is the mean residue weight in Daltons, P is the path length in cm, CONC is the concentration in mg/ml, and 3298 is a specific constant.

Secondary structure content was calculated from CD spectra using Dichroweb server, using the CDSSTR algorithm, with reference set 4 optimised for 190 – 240 nm (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml, Whitmore and Wallace, 2004).

2.2.4.6 Determination of protein solubility

The solubility of Agp1 proteins was measured using a method similar to that described in Trevino et al., (2007). The solubility of the BV adducts was measured instead of that of the apoproteins, because Agp1 apoproteins were found to dimerise via a disulphide bond, which may affect their solubility (Section 6.5.1). Maintaining the monomeric state of apoproteins requires reducing agents, for example DTT, which may interfere

with spectrophotometric measurements. The BV adduct is suspected to be the predominant form of phytochromes in *A. tumefaciens* (Section 6.7.4), therefore characterisation of BV adducts was deemed more physiologically relevant.

7.27 µl of the ammonium sulphate buffer (3.3 M $(NH_4)_2SO_4$, Section 2.2.3.7.2) was added to 12.72 µl of Agp1 holoproteins in SEC buffer (Section 2.2.3.7.6) to obtain 20 ml of holoprotein in 1.2 M $(NH_4)_2SO_4$. Only highly concentrated samples were used to ensure the appearance of precipitation, which is an indication of saturation. The solution was equilibrated for 1 minute at 25 °C in the dark, and centrifuged for 1 minute at 16,000 × g (25 °C, darkness). 10 µl of the supernatant was mixed with 440 µl of the SEC buffer containing 1.2 M $(NH_4)_2SO_4$ for quantifying the protein concentration by absorption spectrophotometry at 280 nm.

2.2.5 Bioinformatics

2.2.5.1 Protein characterisation

Basic biochemical properties of proteins such as molecular weight, pI, and ε280 were estimated by submitting the sequence to the Protein calculator server (http://www.scripps.edu/~cdputnam/protcalc.html).

Domains were identified and annotated from amino acid sequences of proteins using the SMART server (http://smart.embl-heidelberg.de/, Letunic et al., 2012).

2.2.5.2 Sequence alignment

Homologous sequences were aligned with ClustalW 2.0.9 (Larkin et al., 2007). Aligned sequences were viewed with CLC sequence viewer (CLC bio). The colour scheme of amino acid residues is as shown in Fig. 2.2.

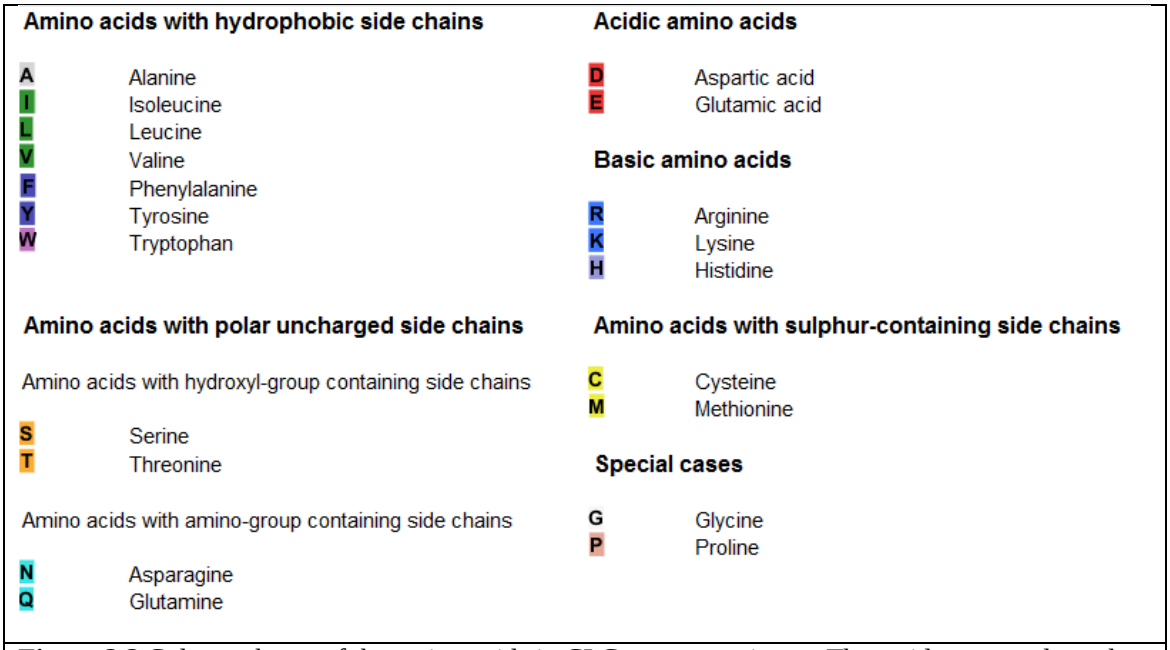


Figure 2.2 Colour scheme of the amino acids in CLC sequence viewer. The residues are coloured according to their chemical properties.

2.2.5.3 Surface entropy reduction server

Clusters of residues best suited for introducing mutations according to "surface entropy reduction" (SER) strategy were chosen using a combination of manual and automated assessments based on criteria described in Goldschmidt et al. (2007). The criteria were as follows.

- The cluster consists of amino acids with high conformational entropy that are likely to occur on the proteins surface and unlikely to participate in protein-protein interaction. Default mutable amino acids are lysine, glutamate, and glutamine.
- Such a cluster must occur in a loop region. SER mutations have been most successful when introduced in loops as they are likely to be solvent exposed. SER mutations were reported to be less effective when introduced on the surface exposed face of a helix.
- The target mutable residues must not be mutated if they are conserved, because conservation implies functional significance of the residue.

Surface exposure of the suggested mutable clusters was checked based on the preliminary protein structure where possible. Where the structural information was absent, a homology model (Section 2.2.5.3) was used instead.

2.2.5.4 Homology modelling

Modeller was used to create homology models using template structure and target protein sequence according to the procedure given in the tutorial at http://salilab.org/modeller/.

2.2.6 Crystallisation of proteins

2.2.6.1 Crystallisation screening

Crystallisation screening was carried out with a Mosquito robot (TTP labtech). For the screening of crystallisation conditions for a new protein, JCSG+, PACT premier, PGA, and Morpheus kits (Molecular Dimensions) were often used for sparse matrix and systematic screening. Reservoir volumes were 80 µl and 45µl for 2-drop and 3-drop 96-well plates, respectively. Drop sizes were 400 nl, containing the protein solution and the reservoir solution at a volume ratio of 1:1. The plates were incubated in the darkness either at 14, 19, or 25°C. Plates were inspected daily under a microscope with a green filter during the first week of setup, and then every 2 weeks for up to 3 months. Drops were scored according to a scheme described in Benvenuti & Mangani, (2007).

2.2.6.2 Optimisation of crystallisation conditions

The optimisation of crystallisation conditions was carried out mostly with 24-well XRL plates with hanging drop or sitting drop set ups. Optimisation parameters included the chemical composition of the reservoir solution, volumes of the reservoir and the drop, the volume ratio of the protein solution and the reservoir solution, protein concentration and temperature. Additive screen 1 (Molecular Dimensions) was used to further optimise the chemical composition of the crystallisation condition. Streak seeding with cat whiskers and solution microseeding was used to grow crystals better suited for analysis (Bergfors, 2009, Chapter 6).

2.2.7 Crystallographic structure determination

2.2.7.1 Diffraction data collection

All protein crystals were vitrified with liquid nitrogen under the presence of a suitable cryoprotectant. The strategy for data collection was determined either manually or automatically by EDNA software available at most MX beamlines at synchrotrons. Experimental parameters depend on a multitude of factors. The wavelength of X-rays

obtained from a rotating copper anode is fixed at 1.54 Å, whereas the wavelength is tuneable at most synchrotron beamlines. The exposure time per diffraction image depends on the intensity of the beam and sensitivity of the detector. Generally, the detector distance was adjusted in order to maximise the pixel resolution of the diffraction image without losing the spots at high diffraction angles. The rotation angle per image was set to avoid the overlap of the spots, and this depends on the quality of the incident X-rays and the mosaicity of the crystal. The total angular range for data collection depends on the space group of the crystal as well as how susceptible the crystal is to radiation damage.

2.2.7.2 Data processing

Either the CCP4 (Bailey, 1994) or the Phenix (Adams et al., 2010) software suite were used for most part of the data processing. Both suites contain a number of identical softwares, and CCP4 was used unless otherwise specified.

2.2.7.2.1 Indexing and scaling

Datasets were indexed with either iMosflm 1.0.5 or 1.0.6 (Leslie and Powell, 2007) or XDS (Kabsch, 2010). The default was iMosflm 1.0.5, but iMosflm 1.0.6 and XDS were used for indexing data from a Pilatus detector. Intensities were scaled with Scala (Winn et al., 2011) and converted to amplitudes with Truncate (French and Wilson, 1978).

Statistics given at the indexing stage in iMosflm can give indications as to which space group is correct. Potential Bravais lattices are listed in order of increasing penalty. Penalty is a value correlated with σ that indicates the degree of mismatch between the positions of the observed reflections and the calculated reflection of the idealised lattice. σ (x,y) and σ (φ) values are the root mean square difference between the observed and calculated reflections in mm and °, respectively (Battye et al., 2011).

Reliability factors were calculated with the following equations (Evans, 2006).

$$R_{\text{merge}} = R_{\text{sym}} = \sum_{\mathbf{h}} \sum_{l} |I_{\mathbf{h}l} - < I_{\mathbf{h}} > | \; / \; \sum_{\mathbf{h}} \sum_{l} < I_{\mathbf{h}} >$$

$$R_{\text{meas}} = R_{\text{r.i.m}} = \sum_{\mathbf{h}} \sqrt{\frac{n_{\mathbf{h}}}{n_{\mathbf{h}} - 1}} \sum_{l} |I_{\mathbf{h}l} - < I_{\mathbf{h}} > | \; / \; \sum_{\mathbf{h}} \sum_{l} < I_{\mathbf{h}} >$$

$$R_{\text{p.i.m}} = \sum_{\mathbf{h}} \sqrt{\frac{1}{n_{\mathbf{h}} - 1}} \sum_{l} |I_{\mathbf{h}l} - < I_{\mathbf{h}} > | \; / \; \sum_{\mathbf{h}} \sum_{l} < I_{\mathbf{h}} >$$

where I is the intensity, and $n_{\mathbf{h}}$ is the number of observations of reflection **h**. $I_{\mathbf{h}l}$ is the lth observation of reflection **h**, and $< I_{\mathbf{h}} >$ is the mean intensity of all reflections which are related by crystallographic symmetry. Reliability factors indicate accuracy of the measured intensities. $R_{\text{merge(sym)}}$, the most widespread statistics, is calculated by dividing the sum of differences between each intensity measurement $I_{\mathbf{h}l}$ and the mean intensity $< I_{\mathbf{h}} >$ by what is essentially the sum of all intensities. In recent years, the statistical incorrectness of $R_{\text{merge(sym)}}$ has been pointed out, because the $R_{\text{merge(sym)}}$ rises as the multiplicity increases. More improved statistics have been introduced in recent years (Blow, 2002). $R_{\text{meas(r.i.m)}}$ is a more robust R_{sym} which indicates the reliability of individual measurement independent of multiplicity (Diederichs and Karplus, 1997). Each reflection's contribution is adjusted by a factor of $\sqrt{n_{\mathbf{h}}}/(n_{\mathbf{h}}\text{-}1)$, where $n_{\mathbf{h}}$ is the multiplicity. The factor of $\sqrt{1}/(n_{\mathbf{h}}\text{-}1)$ found in $R_{\text{p.i.m}}$, which was originally discussed in Weiss & Hilgenfeld (1997), yields information about the average precision of the data.

Truncate program also subjects datasets to "L-test" to test for the presence of twinning (Padilla and Yeates, 2003). This test utilises the L-values calculated from local pairs of reflection intensities, and is insensitive to phenomena such as anisotropic diffraction and pseudo-symmetry. The L-value is calculated with the following equation.

$$L = (I_1 - I_2)/(I_1 + I_2)$$

Where I_1 and I_2 are intensities of a local pair of reflections.

Quality of a diffraction dataset can also be judged from the intensity statistics in the output from the scaling operation. The ratio <I>/<σ(I)> where <I> is the estimated mean intensity and <σ(I)> is the mean standard deviation of the measurements. <I>/<σ(I)> can be plotted as a function of resolution (Blow, 2002).

2.2.7.2.2 Molecular replacement

Matthews software (Kantardjieff and Rupp, 2003) was used to calculate the Matthews coefficient and solvent content from the unit cell parameter and the molecular mass of the protein in the unit cell. Phases were determined by molecular replacement using the known homologous structures with Phaser (McCoy et al., 2007). Some search models

were treated with Chainsaw (Stein, 2008) where non-conserved residues were pruned to the C_β atom of the side chain.

2.2.7.2.3 Model building and refinement

The initial model used for refinement was either the search models used for the molecular replacement or a model generated automatically by AutoBuild module of Phenix (Adams et al., 2010). The former method was always used unless otherwise specified. Structures were modelled and refined by iterative rounds of manual and automatic refinement using Coot (Emsley et al., 2010) and Refmac5 (Murshudov et al., 2011).

Refinement weighting of Refmac5 determines the strength of geometric constraints imposed onto the structure during refinement routines. Smaller weighting value imposes tighter constraint and vice versa. Refinement weighting was mostly set at default and determined automatically by the program, but a small value such as 0.1 was chosen where tighter geometric constrain was preferred.

2.2.7.2.4 Generation of library for biliverdin

The library file necessary for determining the ideal geometry of a ligand that is not included in the default collection of Coot and Refmac5 were generated with the following procedure. The coordinate of the BV molecule was extracted from the high resolution structure of DrBphP (2O9C), and was submitted to the PRODRG server (http://davapc1.bioch.dundee.ac.uk/prodrg/, Schüttelkopf and van Aalten, 2004). The bond lengths, angles, and types were manually modified according to a high resolution crystal structure of biliverdin dimethyl ester (Cambridge structural database accession: bilver) where applicable. The resulting library file was used during refinement with Coot and Refmact5.

2.2.8 Structural analyses

2.2.8.1 Visualisation of molecular models

Protein structural coordinates in PDB format were viewed with PyMOL (DeLano, 2002).

2.2.8.2 Visualisation of protein-ligand and protein-protein interaction

Ligplot 4.5.3 was used to plot protein-ligand and protein-protein interactions (Wallace et al., 1995).

2.2.8.3 Structural superposition

Global protein structure superpositions were achieved by "SSM superpose" function of Coot or *super* command of PyMOL. When the range of residues needed to be specified for local structural superposition, "LSQ superpose" function of Coot was used.

2.2.8.4 Evaluation of surface electrostatic properties

Protein structure coordinate files were subjected to analysis of Poisson-Boltzmann electrostatics calculations with PDB2PQR (Dolinsky et al., 2007), and subsequently the electrostatic surface potential diagrams were generated with Adaptive Poisson-Boltzmann Solver (APBS) plugin for PyMOL (Baker et al., 2001).

APBS electrostatic surface potential diagrams were coloured by electrostatic potential (-5 kT/e, red to +5 kT/e, blue) with the "Colour by potential on sol. acc. surf." option enabled.

2.2.8.5 Secondary structure assignment to protein structures

Secondary structures in a protein structure were objectively assigned with DSSP (Joosten et al., 2011, Kabsch and Sander, 1983). Each residue was assigned to be either α-helix, isolated β-bridge, β-strand, 3/10-helix, 5-helix, hydrogen bonded turn, bend, or no secondary structure. The proportion of the secondary structure relative to either the modelled residues or the total residues in the construct was calculated.

2.2.8.6 Calculation of the angle between two helices

Intersecting angle between two α-helices was calculated with QHELIX (Lee et al., 2007). Chou algorithm was selected unless otherwise specified. Ten terminal residues were selected in order to calculate the angle between the termini of a continuous helix, as selection of limited segments (e.g. five residues) led to aberrant results with large disagreement between Chou and Kahn algorithms.

2.2.8.7 Protein structure surface area calculation

Solvent accessible area and contact area of protein structures were calculated by Areaimol (Saff and Kuijlaars, 1997). The dimer interface area was calculated from the difference in solvent accessible surface area per chain in a monomeric and dimeric form of a molecule.

2.2.8.8 Calculation of the angle between planes

The averaging plane was calculated form selection of atoms by Olex2 (Dolomanov et al., 2009) with the *mpln* command. The angle between two planes was calculated by the *sel* command. The coordinate file format had been converted form .pdb to .xyz by Discovery studio 3.1 (Accelrys).

Chapter 3: Structural characterisation of Agp1 in the Pr-form

3.1 Overview

Protein structures at atomic detail combined with available biochemical and biophysical data can provide deep insights into a protein's functional mechanism. As a first step of gaining structural understanding about the functional mechanisms of phytochromes, crystallisation of Agp1 in the Pr-form was attempted.

3.2 M20: Chromophore binding domain (PAS-GAF)

To solve the structure of Agp1 in the Pr-form, a relatively short protein product termed M20 was used first (Noack et al., 2007). This construct encompasses the chromophore binding domain (CBD) which includes the PAS and GAF domains, and is the minimum domains necessary for lyase function that is responsible for covalent attachment of the bilin chromophore. In classical bacteriophytochromes, chromophore attachment to the CBD results in formation of the Pr-form (Noack et al., 2007, Wagner et al., 2005). There are phylogenetically distinct phytochromes, however, that only need a GAF domain to form a functional holoprotein that undergoes photoconversion from the Pr ground state into the Pfr-form (Cornilescu et al., 2008).

M20 was initially considered as a suitable target for structural elucidation of Agp1 for the following reasons. Firstly, the rate of successful crystallisation as defined by a statistical study from a structural genomic study is higher for the protein with the size of M20 (313 residues) than M15 (504 residues) (Slabinski et al., 2007). Secondly, though M20 and equivalent constructs of other phytochromes are capable of forming a covalent bond with the bilin followed by an emergence of increased absorption in the Q-band, they are not able to form the Pfr-form neither by photoconversion nor by dark-reversion in case of bathy-phytochrome (Noack et al., 2007, Wagner et al., 2005, Yang et al., 2008, Yang et al., 2007). The inability to form the Pfr-form was regarded as an advantageous trait for crystallisation trials as it increases the chance of protein being more conformationally homogeneous, a favourable characteristic for crystallisation (Benvenuti and Mangani, 2007). Despite the inability for M20 holoprotein to undergo full photoconversion, M20 holoprotein was always handled under the safelight, because some spectral changes are evident when M20 holoprotein was exposed to the light which may be associated with conformational change (Noack et al., 2007). The choice of such a shorter construct for crystallographic structure determination may be justified

by the fact that the currently available phytochrome structure with the highest resolution is of the PAS-GAF domains of DrBphP (Auldridge et al., 2012). Finally, the structural information of a shorter construct may help to solve the structure of a longer construct, as it may be used as a search model in molecular replacement during phase determination.

3.2.1 Expression and purification of M20-BV protein

The gene of the M20 protein was expressed as described in Section 2.2.3. Typically 4 litres of LB-ampicillin media were used per batch to express and purify protein. The purity of M20 apoprotein following affinity chromatography was checked with SDS-PAGE (Fig. 3.1).

Purified apoprotein was assembled *in vitro* with BV for two hours in the darkness. IEX was applied to separate holoprotein from apoprotein as it became evident during an analytical IEX that BV incorporation is less efficient with M20 apoprotein than with M15, which is likely due to the lack of the PHY domain. The fractions containing the holoprotein were pooled, concentrated, and applied to SEC to isolate a conformationally homogeneous population of the protein, as well as to exchange the buffer to SEC buffer (20 mM Tris-HCl pH 7.8 at 20°C, 1 mM EDTA with or without 50 mM NaCl). Purified protein was concentrated to typically 20 – 30 mg/ml, and either used immediately or flash-frozen with liquid nitrogen in 0.1-1.0 ml aliquots and stored at -80 °C in darkness.

Purified M20-BV had a typical SAR value (Section 1.3.3) of 1.2–1.7 (Fig. 3.3), which on average is higher than that of M15-BV (1.23, Scheerer et al., 2006). This is still lower than the expected range of SAR values which can be estimated based on the simple assumption that while εQ_{max} of M15-BV and M20-BV are the same, M20-BV will have a lower ε280 due to smaller number of amino acids with high absorption coefficient. Based on that assumption, the SAR value of M20-BV should accord to the equation given below.

$$SAR_{M15} \times \varepsilon 280_{M15} \div \varepsilon 280_{M20} = 1.23 \times 77950 \div 34280 = 2.80$$

The experimentally determined SAR value for M20-BV is lower than the calculated SAR, which may be attributed to either persisting apoprotein contamination and/or due to the lack of the PHY domain. The assembly of phytochromes with the bilin chromophore is characterised by an increase in absorption in the Q-band, and it has been observed in phytochrome structures that the configuration of the bilins are

conserved, including its characteristic tilt of the D-ring with respect to the co-planar B- and C-rings (Wagner et al., 2005). Such a spatial constraint to the BV in the holoprotein which is not found in the free bilin in solution may be correlated with a high extinction coefficient and λ_{max} in the Q-band in holophytochromes. The absence of the PHY domain in M20 could lead to structural differences of the bilin from that in the M15 structure which are responsible for the different absorbances in the Q-band. Increase in the absorption in Q-band in context of bilin configuration and chromophore-protein interaction is not well documented in the literature owing to its complexity. A relevant observation in this context could be that there is an increased absorption in the Q-band of Agp2 holoprotein with 5Ea15Ea-BV chromophore, which was assumed to be the result of improved protein folding following NAP column separation (Inomata et al., 2009). Also, it was found in a study with Agp2 that the PCM fragment is unable to form holoprotein with normal spectral properties, and the absorption spectrum of the CBD fragment is affected even further (Karniol et al., 2005). A high tilt of the D-ring of the bilin in CBD structures of DrBphPs were attributed to the lack of the PHY domain (Essen et al., 2008). These findings demonstrate that not only each domain contributes to bilin incorporation, but the degree of contribution differs among phytochromes, despite similar domain structures.

The assembly mixture was subjected to purification with IEX, which successfully separated the holoprotein from the apoprotein (Fig. 3.2). Typically between 20–40% of the total eluate was apoprotein despite BV being supplied in molar excess during assembly. Such inefficient uptake of BV was not observed in case of M15 (Section 6.2.1), and the inefficiency of M20 is attributed to the lack of the PHY domain, even though to date not much is known about its role in bilin incorporation.

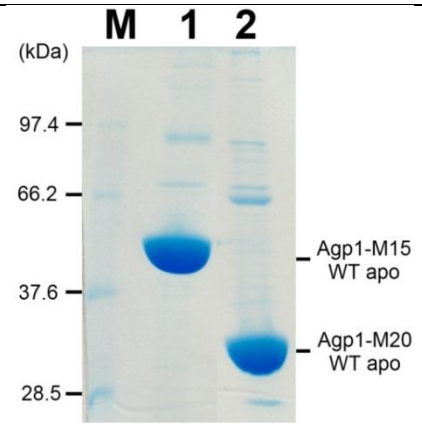


Figure 3.1 SDS-PAGE showing affinity purified M20 apoprotein. Lane M, 5 µl of Applichem molecular weight marker; Lane 1, M15 apoprotein for comparison; Lane 2, M20 apoprotein. 20 µg of sample was loaded per well in lanes 1 and 2.

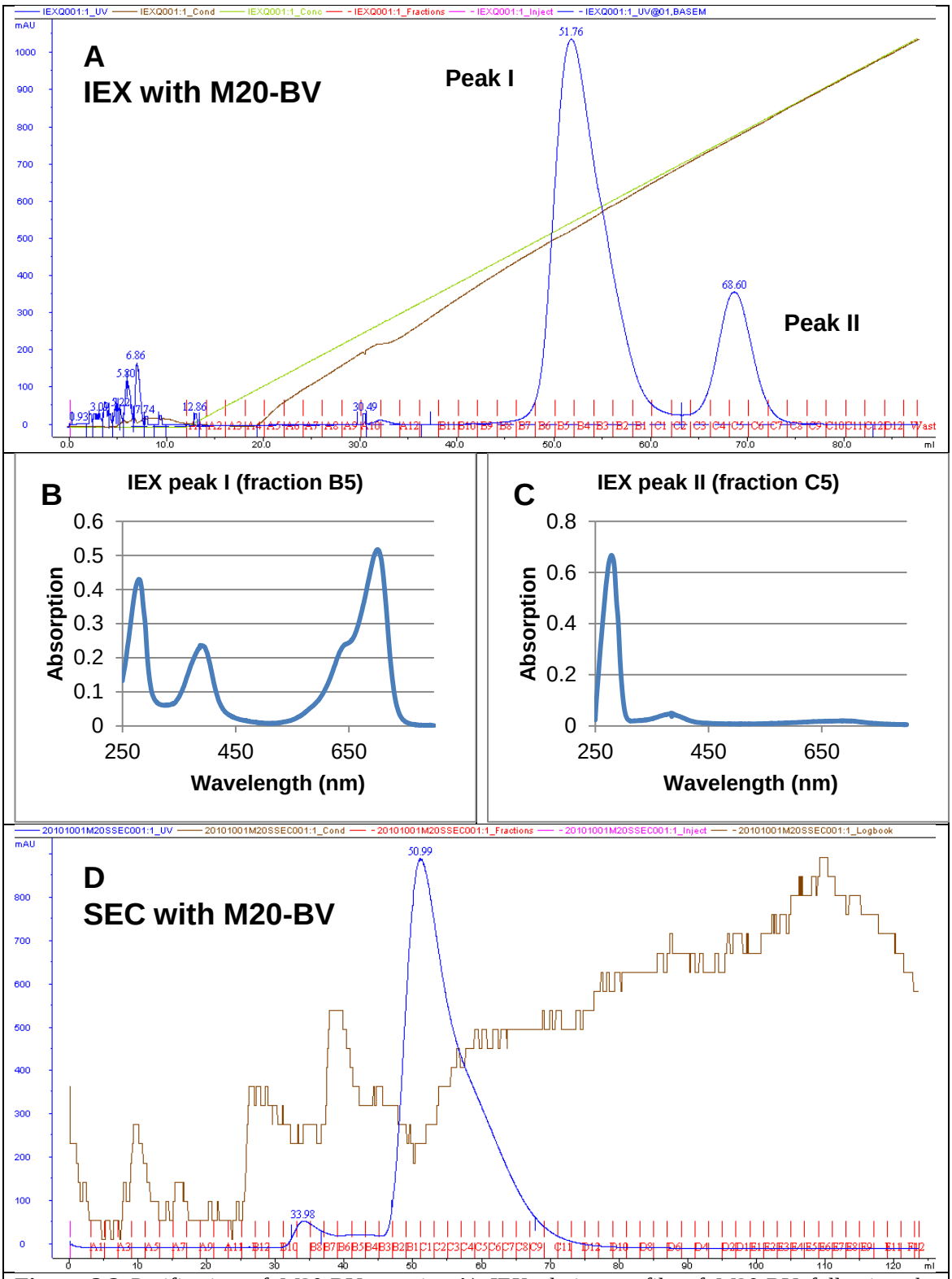


Figure 3.2 Purification of M20-BV protein. A) IEX elution profile of M20-BV following the assembly. UV-vis absorption spectra of the peak I B) and II C) from IEX. The first peak contains M20-BV holoprotein while the second peak consists predominantly of apoprotein. D) A SEC elution profile of M20-BV with a Sephacryl 200 column. The blue, brown, and green lines indicate the absorption (mAU at 280 nm), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively.

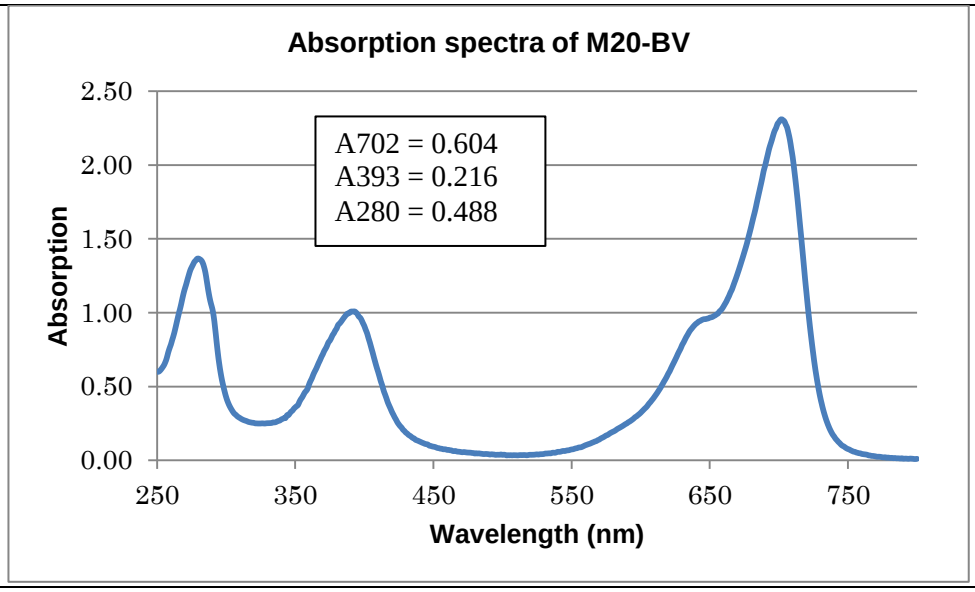


Figure 3.3 UV-vis absorption spectrum of purified M20-BV holoprotein. The SAR value (A702/A280) is 1.69.

3.2.2 Crystallisation screening with purified M20-BV protein

The M20-BV holoprotein was subjected to crystallisation trials using the following screening kits - JCSG+, PACT premier and PGA (Molecular Dimensions), Hampton screen 1 and 2 (Hampton), and JBScreen Classic 1 to 10, JBScreen 1 to 4, JBScreen Kinase 1 to 4 and JBScreen PEG/Salt 1 to 4 (Jena Biosciences). Despite this being the most intensive screening in the present study, no lead condition was identified from it.

M20-BV crystals have been found however, in a preliminary trial in which the solubility of M20-BV holoprotein was studied as a function of precipitant concentration of several types (Fig. 3.4). These crystals were found by chance about 550 days after the plate was set up, long after the period for routine checks had ended. The plate had been exposed to fluctuating temperature and light conditions during the last several months of incubation. The emergence of crystals after a long period is explained by slow evaporation due to imperfect sealing of the crystallisation set-up. In that case the lack of a lead condition in the screening may be rationalised that insufficient protein and/or precipitant concentration were being addressed. This is consistent with a high solubility of M20-BV protein, such that the protein could be concentrated up to 60 mg/ml.

Hexagonal crystals were found in a condition of 5% PEG4000 following about 550 days of incubation (Fig. 3.4.B) were cryoprotected with 20% (v/v) glycerol in the reservoir solution prior to vitrification in liquid nitrogen.

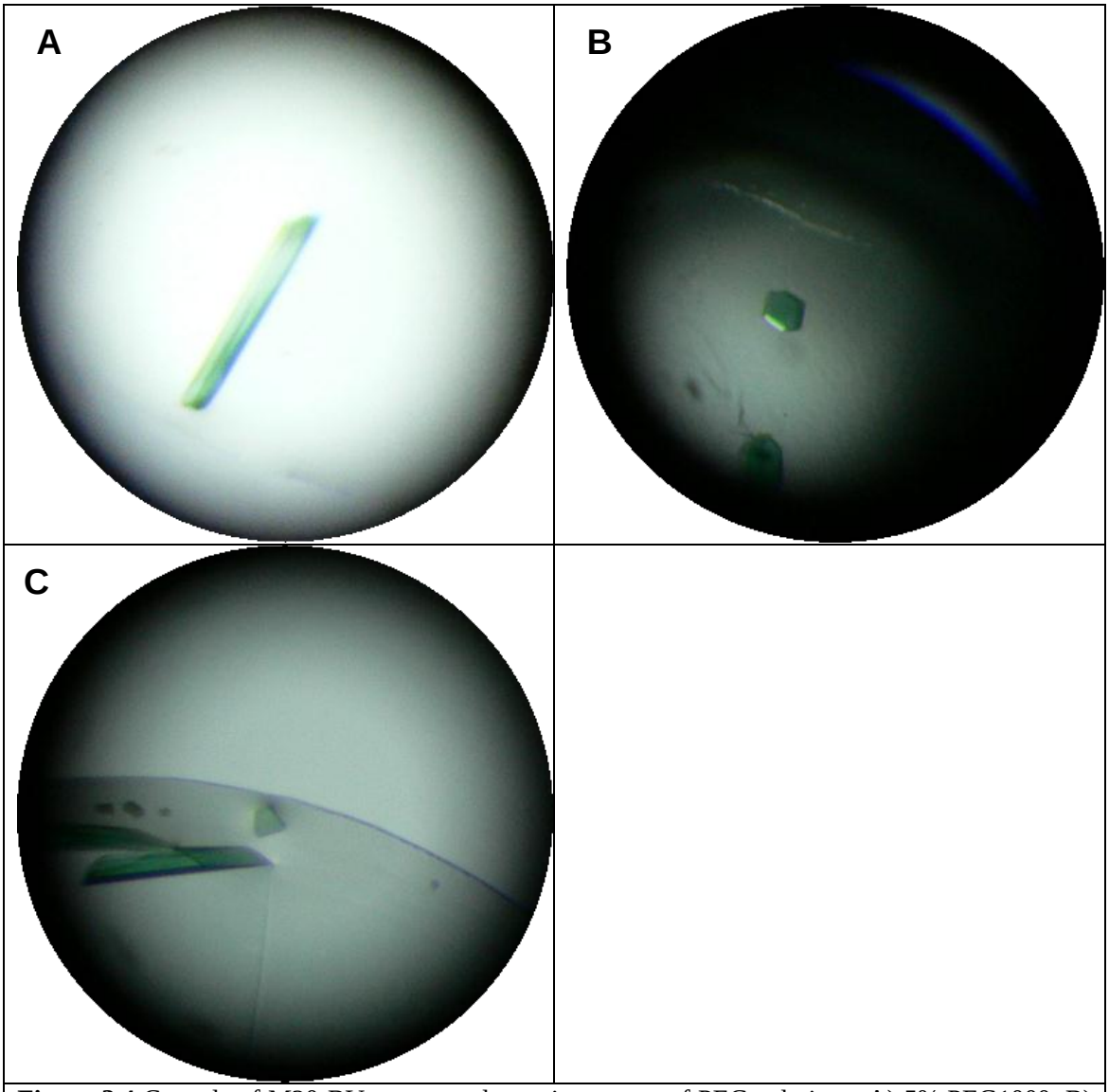


Figure 3.4 Crystals of M20-BV grown under various types of PEG solutions. A) 5% PEG1000, B) 5% PEG 4000, C) 5% PEG8000.

3.2.3 Analysis of M20-BV crystals

Diffraction images were collected for some of the well-formed M20-BV crystals at beamline ID23-1 (ESRF, Grenoble, France). The crystals were judged to consist of M20-BV protein rather than salt, from their green colour and the diffraction patterns, but none of the collected diffraction images was of sufficient quality for the determination of M20-BV structure (Fig. 3.5). Indexing was not possible as the number of reflexions per image was less than optimal for iMosflm which is 200 – 1000 spots per image.

Low diffraction quality may be attributed to any of the factors such as inherently low diffraction quality of M20-BV crystals, conformational heterogeneity by exposure to the light, and chemical heterogeneity as a result of protein degradation during the prolonged incubation period.

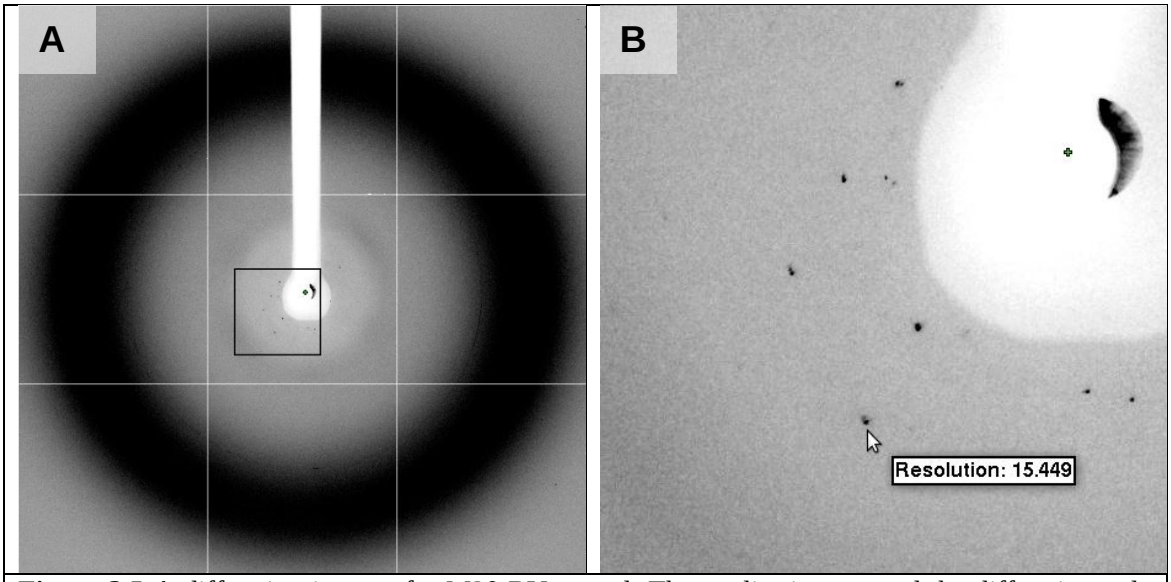


Figure 3.5 A diffraction image of a M20-BV crystal. The quality is poor and the diffraction only occurs maximally to a resolution of about 15.5 Å.

3.3 M20s: truncated M20 construct

After the initial attempts to solve the structure of M20-BV have failed due to the inability to obtain crystals with suitable diffraction quality, the gene construct was modified in hope that the produced protein may crystallise more readily. The modification corresponding to the truncation of the first nine amino acid residues at the N-terminus was considered for the two separate lines of reasoning.

Previously a variation of Agp1 M15-BV without the first N-terminal nine amino acids has led to the formation of crystals with better diffraction quality (P. Scheerer, personal communication). Therefore the omission of this region was proposed to result in the successful crystallisation of M20 protein.

Secondly, as described in Section 1.3.7.1, whether the first nine amino acids at the N-terminus is part of the endogenously expressed protein in *A. tumefaciens* is unclear. Omission of the N-terminal region has been shown to have little effect on the biochemical activities of Agp1 M15-BV protein *in vitro* (Noack et al., 2007).

Optimum truncation of a protein could result in a protein with preferable solubility and crystallisation propensity. The optimum truncation point may be determined by creating a library of differently truncated proteins (Listwan et al., 2009). For the purpose of this project, however, such an extensive optimisation was not carried out on M20 protein in favour of solving the structure of a longer construct, M15, the structure of which will be more informative than M20.

3.3.1 Generation of M20s plasmid construct

An expression construct termed pNM19a bearing the coding sequence of M15 Δ1-9 was selected as the basis for generating an expression vector for M20 Δ1-9, therefore generation of the construct only involves the omission of the sequence of the PHY domain. M20 Δ1-9 construct will be referred to as M20s hereafter. Also a construct based on pNM19a has the benefits of pET based vector system which allows controllable, high levels of expression under compatible host cells. Details of the generation of M20s plasmid construct is shown in Appendix 2.

3.3.2 Expression and purification of M20s-BV protein

M20s apoprotein was expressed, purified and assembled following the identical protocol as for the M20 protein, resulting in holoprotein with absorption spectra indistinguishable from that of the M20-BV protein. The protein was concentrated to typically 20 – 30 mg/ml prior to screening for crystallisation conditions.

3.3.3 Crystallisation screening with purified M20s-BV protein

Conditions from JCSG+, PACT premier and PGA kits were screened to crystallise M20s-BV holoprotein. Screening was carried out following the identical protocol as for the M20-BV holoprotein (Section 3.2.2). Crystallisation trials were also carried out around the condition in which M15-BV crystallise well (200 mM $CaCl_2$, 20% (w/v) PEG3350, 100 mM Tris-HCl pH 8.0 at 20 °C), but without success. Promotion of crystal formation was attempted by cross-seeding with Agp1 M15-BV crystals into pre-equilibrated drops of M20s-BV using cat whiskers. No lead crystallisation condition has been identified in any of the trials, however.

3.4 M15 and M15s: Photosensory core module (PAS-GAF-PHY)

On the contrary to the initial anticipation, neither M20-BV nor M20s-BV protein crystallised readily to allow for the structural determination. Hence the focus was turned to M15 construct and its variation, which encompass the entire photosensory core module (PCM) including the PAS, GAF, and PHY domains. Due to the presence of the PHY domain, the PCM of phytochromes are capable of undergoing photoconversion from the Pr-from into the Pfr-form (Noack et al., 2007). Structural information of phytochrome PCM will therefore give deeper insights for the functionalities of phytochromes, such as the mechanism of photoconversion and signal transduction, compared to that of CBD.

To this end, three versions of the wild type (WT) M15 were attempted for crystallisation. The differences in the coding sequence and the strains of *E. coli* cells used for protein expression are listed in Table 3.1. Crystallisation of M15 protein lacking the first nine N-terminal residues was also attempted as it was observed to produce better crystals (P. Scheerer, personal communication). This shorter construct without the nine N-terminal residues is referred to as M15s hereafter, while proteins expressed from both pAg1-M15 and pNM9a are generally referred to as M15, unless otherwise specified.

Throughout the present study, only proteins with the chromophore-binding cysteine at the 20^{th} position were used for crystallisation purposes to ensure the formation of covalent bond between the protein and the bilin moiety. For some biochemical studies, however, a mutant that cannot form a covalent bond with a BV was used. The protein produced from a construct termed pNM9c had the cysteine residue at 20^{th} position replaced with an alanine.

Construct	Base plasmid and source	Sequence description	*E. coli* strain used for protein expression
pAg1-M15	pQE12 (Qiagen)	Agp1 1 – 504 and hexahistidine tag	XL1-Blue
pNM9a	pET21b (Novagen)	Agp1 1-504, LE (XhoI site), and hexahistidine tag	BL21(DE3)
pNM9c	pET21b (Novagen)	Agp1 1-504 with C20A mutation, LE (XhoI site), and hexahistidine tag	BL21(DE3)
pNM19a	pET21b (Novagen)	Agp1 10-504, LE (XhoI site), and hexahistidine tag	BL21(DE3)

Table 3.1 List of M15 constructs at the start of the project. Proteins with the WT sequence were used for crystallization trials.

3.4.1 Expression and purification of M15 / M15s-BV protein

M15 and M15s apoproteins were overproduced as described in Section 2.2.3.6.2, and the purification protocol was identical regardless of the expressed protein variation or the strains of cells used, including the C20A mutant of M15. Typically 4 litres of LB-ampicillin media were used per batch to express and purify the protein. Purity of M15(s) apoprotein purified by affinity chromatography was checked with SDS-PAGE (Fig. 3.6).

Purified apoprotein was assembled *in vitro* with BV. Assembled M15(s)-BV holoprotein was purified by SEC with +/- 50 mM NaCl, 20 mM Tris-HCl pH 7.8 at 20°C, and 1 mM EDTA using either Sephacryl 200 16/60 or Superdex 200 10/300 column (GE Healthcare) (Fig. 3.7). Holoprotein containing fractions with high SAR values were pooled and concentrated to 20-30 mg/ml. Absorption spectra of the purified protein was measured in order to determine the protein concentration and to inspect the SAR value to infer the relative purity of the holoprotein (Fig. 3.8). Purified protein was either used immediately or flash-frozen with liquid nitrogen in 0.1-1.0 ml aliquots and stored at -80 °C in darkness.

Additional purification step to purify the holoprotein from the apoprotein was deemed unnecessary for M15(s)-BV, as judged from analytical IEX in Section 6.2.1. As far as M15 fragment is concerned, the difference in elution volumes in IEX between apoprotein and the BV adduct is comparatively higher than the difference between the apoprotein and the 15Ea adduct. IEX was routinely used to purify the 15Ea adduct from the apoprotein, therefore IEX is in theory capable of isolating the BV adduct from the apoprotein. Yet no apoprotein peak was ever identified when M15-BV holoprotein was subjected to IEX, and this was attributed to more efficient BV assembly by M15 than M20, possibly by the virtue of the PHY domain.

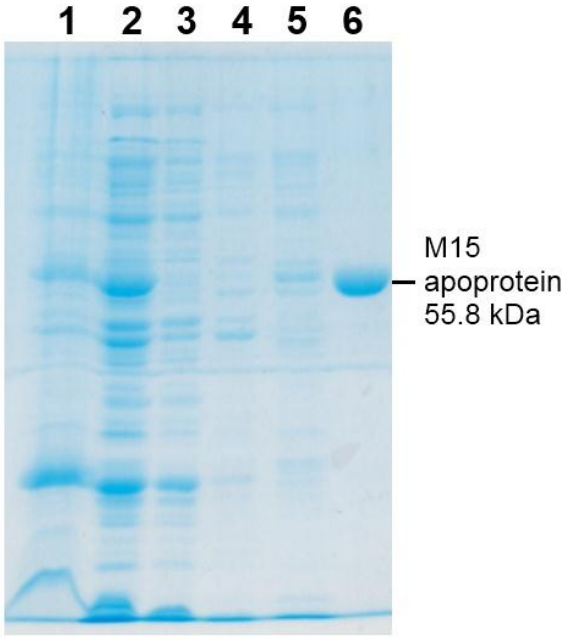


Figure 3.6 SDS-PAGE gel showing the expression and purification of M15 apoprotein. Lane 1, insoluble fraction after cell lysis; Lane 2, clear lysate after cell lysis; Lane 3, flow-through from the affinity column following sample application; Lane 4, flow-through from the affinity column during washing stage, the first half; Lane 5, flow-through from the affinity column during washing stage, the second half, Lane 6, purified apoprotein.

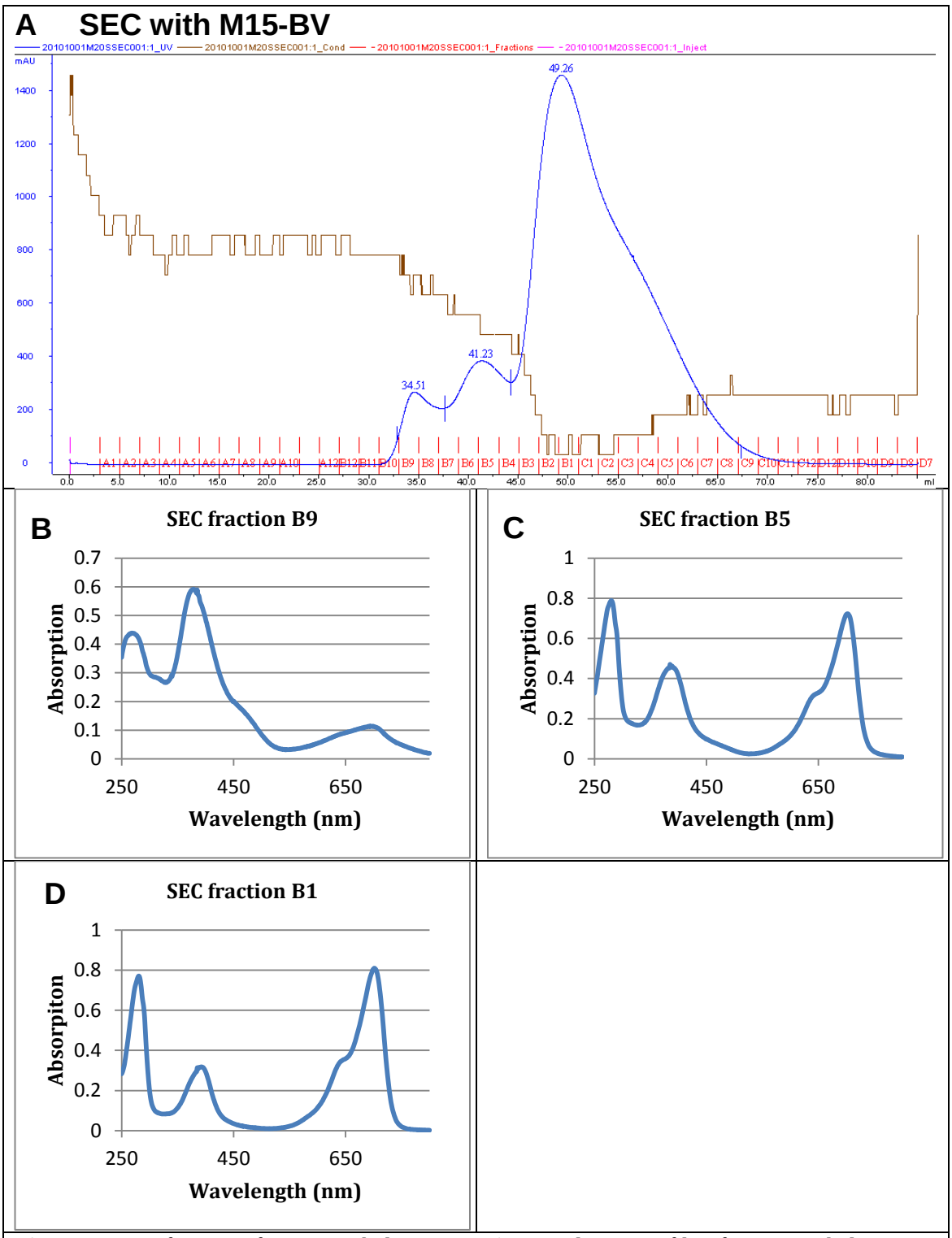


Figure 3.7 Purification of M15-BV holoprotein. A) SEC elution profile of M15-BV holoprotein. Blue and brown lines indicate the absorption (mAU at 280 nm) and conductivity (mS/cm), respectively. UV-vis absorption spectra of peak-containing fractions from SEC. B) Fraction B9, C) Fraction B5, D) Fraction B1. Fraction B9 was assumed to be improperly folded protein associated with BV. Similar elution profile was observed between Sephacryl 200 and Superdex 200. Of the peaks with normal spectrophotometric properties, the minor (fraction B5) and the major peak (fraction B1) were found to correspond to the dimer and the monomer, respectively (Appendix 5).

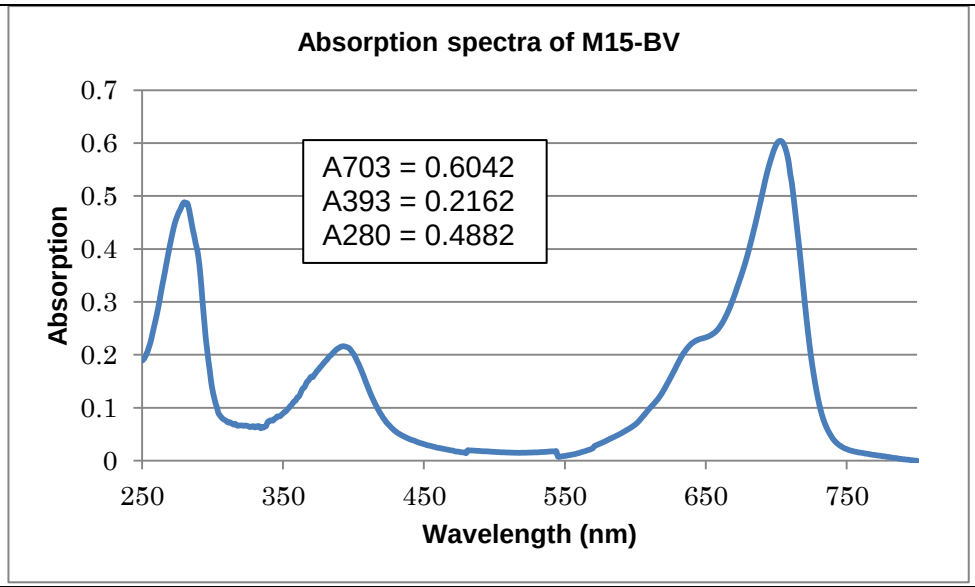


Figure 3.8 UV-vis absorption spectra of purified M15-BV holoprotein. The SAR value is 1.24. Typical SAR value of purified M15-BV samples was in a range of 1.0-1.25. $A_{soret\ (393)}$/A280 is 0.44.

3.4.2 Crystallisation screening with purified M15/M15s-BV protein

Purified M15-BV protein was subjected to crystallisation trials with the following crystallisation screens; JCSG+, PACT premier, PGA, MemStart, and MemSys (Molecular Dimensions), and JBScreen Classic 1 to 10, JBScreen 1 to 4, JBScreen Kinase 1 to 4 and JBScreen PEG/Salt 1 to 4 (Jena Biosciences).

Several lead crystallisation conditions were identified, out of which condition B6 of JBScreen PEG/Salt 1 was chosen to carry out an optimisation, on the basis of favourable crystal morphology of singly growing needles. Optimisation resulted in crystallisation condition (16% w/v PEG3350, 200 mM $CaCl_2$, 100 mM Tris-HCl pH 8.0, 2.5% (v/v) glycerol) in which rod shaped crystals appeared reproducibly, typically within 2 weeks of incubation at 19-25°C (Fig. 3.9). The crystallisation condition reported in Scheerer et al., 2006 (23% w/v PEG3350, 100 mM Tris-HCl pH 8.0, 50 mM NaCl) was also attempted, but it resulted in a crystal form with less volume per length which was deemed less suitable for crystallographic analysis. The two conditions are similar and the only substantial difference is the valence of the cation. It was observed during the optimisation process that conditions with divalent cations such as Mg^{2+} and Ca^{2+} often resulted in crystals with more favourable morphology than those with monovalent cations such as Na^{+} and K^{+}. The type of anion in Ca^{2+} salts did not influence the crystallisation as far as $CaCl_2$, $Ca(NO_3)_2$, and $Ca(C_2H_3OO)_2$ were tested. Crystals with typical dimensions of ≈ 0.15 mm ×0.15 mm × 0.80 mm were

cryoprotected in two steps with the reservoir solutions containing glycerol at 10% (v/v) increments (20% v/v final) prior to the vitrification in liquid nitrogen.

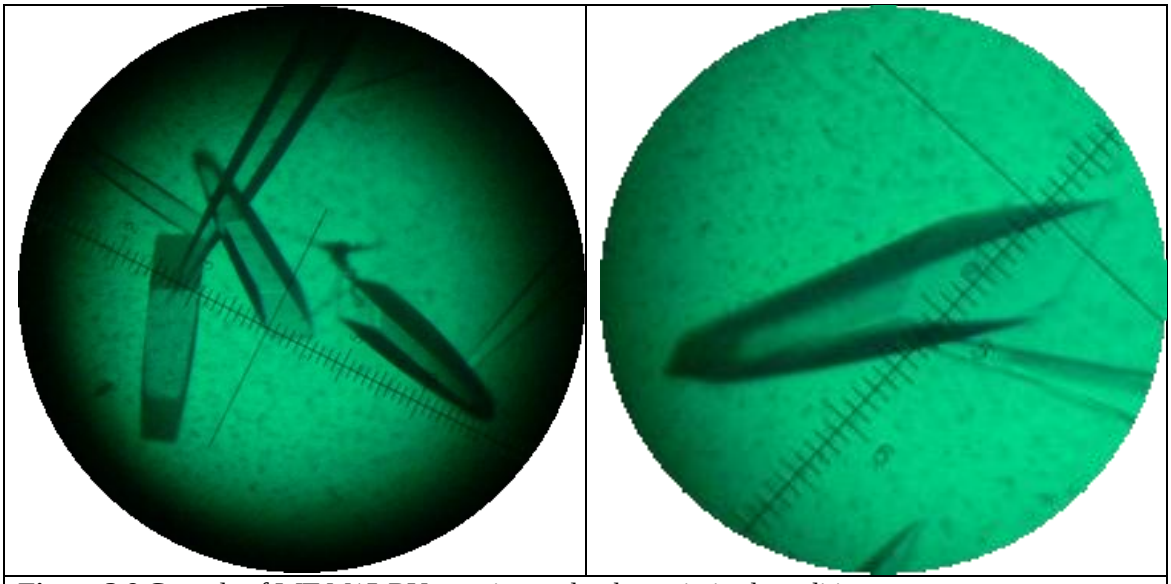


Figure 3.9 Crystals of WT M15-BV growing under the optimised condition.

3.4.3 Analysis of M15(s)-BV crystals

M15(s) -BV crystals were tested for diffraction at the ESRF synchrotron at the following beamlines; ID14-1, ID14-4, and ID23-1. Collected diffraction datasets were analysed for their quality based on statistics such as maximum resolution, signal-to-noise ratio of intensities as expressed by the mean of I/σ (Section 2.2.7.2.1), presence of twinning, and accuracy of intensities as judged by R_{merge}. The best dataset termed 11-2 was collected at beamline ID14-4 at the ESRF synchrotron and it exhibits anisotropic diffraction that extends maximally to 2.77 Å resolution. This was a slight improvement compared to the crystals described in the previous report (Scheerer et al., 2006). The improvement may be attributed to improvement in protein preparation, crystallisation condition, and/or cryoprotection.

Datasets were indexed and integrated with iMosflm, and the intensities were scaled with Scala. The space group suggested by Pointless, I422, was in agreement with the previous report (Scheerer et al., 2006), and the unit cell parameters were essentially the same. The phases were determined by molecular replacement. The search models were either the preliminary structure of the Agp1 M15-BV which had been modelled from a previously collected dataset (Scheerer et al., 2006), or the combination of the high resolution structure of DrBphP CBD (2O9C) and the PHY domain of Cph1 (2VEA). The PHY domain of Cph1 in the Pr-form was judged to be a better search model than the structure of PaBphP in the Pfr-form, based on a comparison with the preliminary structure of M15-BV. The structures of DrBphP and Cph1 were processed with Chainsaw (Stein, 2008) by pruning the side chain atoms to γ-carbon for the non-identical residues.

The refinement proved to be difficult. The difficulties during the refinement of the protein moiety were characterised by the R_{free} value remaining high at about 40%, and weakly defined electron density map in many locations including the bilin binding pocket (Fig. 3.10). A “pathological crystal form” (Dauter et al., 2005) was suspected, and hypotheses were made to identify the source of difficulty in the refinement. Firstly, a case of pseudosymmetry could have led to overestimation of the symmetry in the crystal. Secondly, the problem may be due to a twinning of the crystal. Finally, the problem may be due to the presence of conformational heterogeneity in the dimer. Rather than two chemically identical monomers being related by crystallographic symmetry, monomers may be in a different conformation, and the heterogeneity may

occur either at the protein moiety like in the case of WT PaBphP structure (3C2W, Yang et al., 2008) with four independent dimers in the asymmetric unit or in the vicinity of the bilin binding pocket as reported by the presence of two Cph1 isoforms in its ground state (Song et al., 2011).

The problem of pseudosymmetry may be addressed by processing data in space groups with lower symmetry. To this end, the dataset 11-2 was indexed and scaled with several other candidate space groups suggested by iMosflm (Table 3.3) and the results are shown in Table 3.4. The lowest overall R_{merge} value was observed with P1, but this may be the result of $I_{hl} - \langle I_{\mathbf{h}} \rangle$ of a reflection being added more times in higher symmetry space groups, therefore R_{merge} cannot be used as a reliable criterion to compare the results of scaling between different space groups (Blow, 2002), and for such purpose, multiplicity-weighted reliability factors such as R_{meas} or $R_{p.i.m.}$ should be used (Evans, 2006). Space group P1 exhibited the lowest overall R_{meas} value of 0.060, whereas I422 exhibited the most favourable overall $R_{p.i.m.}$ and mean I/σ of 0.027 and 17.3, respectively. Collectively, identification of the correct space group from the scaling statistics was judged to be difficult, and there is no clear evidence that I422 is the wrong space group.

The presence of twinning in the dataset 11-2 was detected using the L-test (Fig. 3.11). This adds to the complication of identifying the correct space group, as no twinning is possible in I422, as well as reducing the reliability of the final data following the de-twinning operation. Further attempts of structural refinement of WT M15-BV are described later in Section 3.9, where the refined structure of "SER26-BV", a mutated M15-BV was used as the search model. It is noted that the estimated B-value of the dataset is unusually high (94.5 $Å^2$). Since structural heterogeneity has been reported in the PaBphP crystal structure (Yang et al., 2008) and the average B-value and the estimated B-factor are also very high (127 $Å^2$ and 70 $Å^2$, respectively), it is possible that the degree of disorder of the molecules in this crystal form is high, thus posing difficulty over the satisfactory refinement.

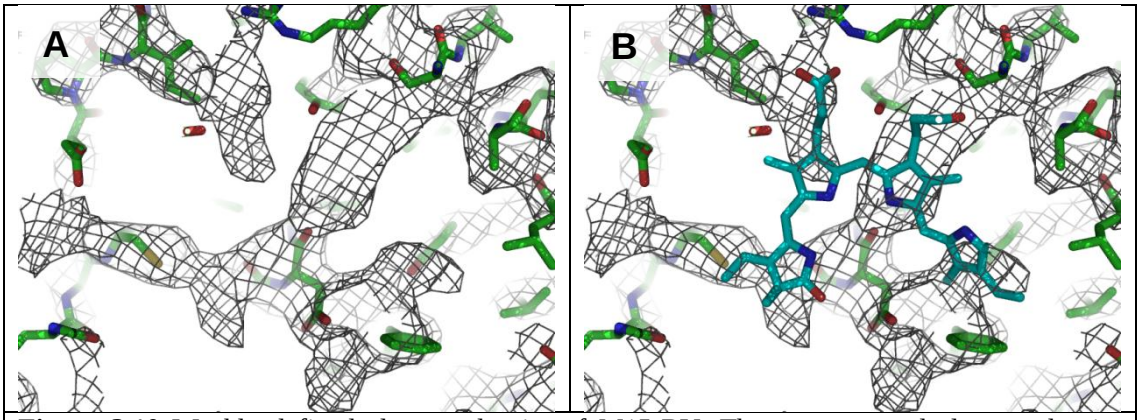


Figure 3.10 Weakly defined electron density of M15-BV. The structure and electron density following molecular replacement with a preliminary model of M15-BV, following refinement with Refmac5. A) Coordinates of protein moiety (green stick) and the 2Fo – Fc map (grey mesh) at the bilin binding pocket. The electron density is contoured at 1.0 σ. B) Same as view as A, but with the BV molecule of the superimposed structure of DrBphP (cyan stick, 2O9C). R/R_{free} values at this stage of refinement are 0.4004/0.4600.

Protein (Dataset)	**M15-BV (11-2)**
Beamline / wavelength [Å]	ID14-4 / 0.9375
Space group	I422
Unit cell dimensions	$a = b = 174.20$ Å, $c = 79.28$ Å $\alpha = \beta = \gamma = 90$ °
Resolution range (outer shell) [Å]	72.16 – 2.77 (2.92 – 2.77)
No. of observations	144322
No. of unique reflections	15690
Multiplicity	9.2
Completeness overall (outer shell) [%]	99.4 (99.8)
<I/σ(I)> overall (outer shell)	17.3 (4.0)
Rmerge overall (outer shell)	0.076 (0.586)
Rmeas overall (outer shell)	0.080 (0.620)
Rpim overall (outer shell)	0.027 (0.200)
Mosaicity (°)	0.49
Estimated B-value (Wilson plot) [$Å^2$]	94.5

Table 3.2 Data-collection and processing statistics of M15-BV dataset 11-2. Values are from data processing in the space group I422.

Number	Space group	Penalty	a	b	c	α	β	γ	σ (x,y)	σ (φ)
1	P1	0	79.4	129.4	129.6	84.5	72.2	72.2	0.12	0.58
2	C2	1	191.8	174.1	79.4	90.0	114.4	90.0	0.12	0.58
3	C2	1	246.5	79.4	1742.0	90.0	134.9	90.0	0.14	0.56
4	I222	2	79.4	174.1	174.7	90.0	90.0	90.0	0.12	0.58
5	C2	2	79.4	246.3	129.6	90.0	107.8	90.0	0.17	0.56
6	C2	2	79.4	246.9	129.4	90.0	107.8	90.0	0.12	0.58
7	F222	4	79.4	246.3	246.7	90.0	90.0	90.0	0.18	0.54
8	I4	5	174.4	174.4	79.4	90.0	90.0	90.0	0.14	0.58

Table 3.3 Results from indexing of dataset 11-2, collected from a M15-BV crystal. Potentially correct space groups and their unit cell parameters are output by analysing a dataset with iMosflm.

	Overall	**Inner**	**Outer**
#1 P1			
R_{merge}	**0.042**	0.024	0.440
R_{meas}	**0.060**	0.035	0.623
R_{pim}	0.042	0.024	0.440
Mean ((I)/sd(I))	9.9	25.3	1.7
#4 I222			
R_{merge}	0.066	0.039	0.546
R_{meas}	0.074	0.042	0.632
R_{pim}	0.033	0.020	0.275
Mean ((I)/sd(I))	13.6	25.7	2.9
#6 C2			
R_{merge}	0.063	0.033	0.527
R_{meas}	0.076	0.042	0.632
R_{pim}	0.042	0.024	0.345
Mean ((I)/sd(I))	9.7	21.0	2.0
#8-1 I4			
R_{merge}	0.066	0.039	0.546
R_{meas}	0.074	0.042	0.632
R_{pim}	0.033	0.020	0.275
Mean ((I)/sd(I))	13.6	25.7	2.9
#8-2 I422			
R_{merge}	0.076	0.045	0.586
R_{meas}	0.080	0.049	0.620
R_{pim}	**0.027**	0.019	0.200
Mean ((I)/sd(I))	**17.3**	30.6	4.0

Table 3.4 Scaling statistics of dataset 11-2 in several space groups. Most favourable values for the overall resolution are emphasized in block.

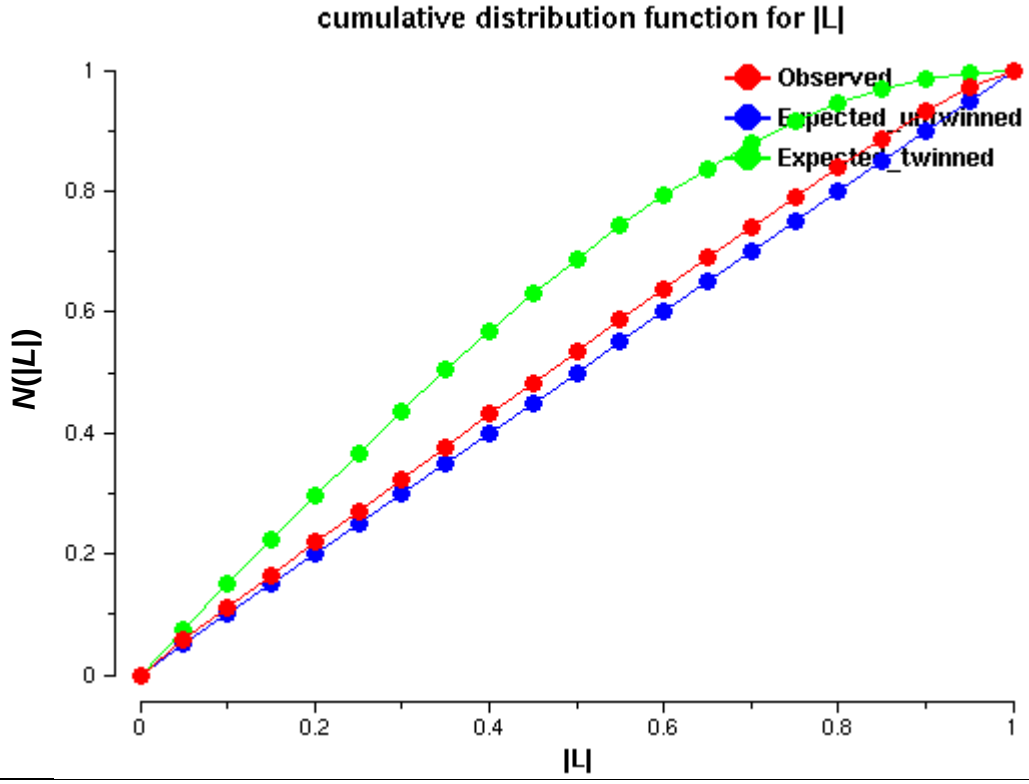


Figure 3.11 L-test for twinning of the dataset 11-2 of M15-BV crystal. Twin fraction was lower than for other datasets of M15-BV crystals.

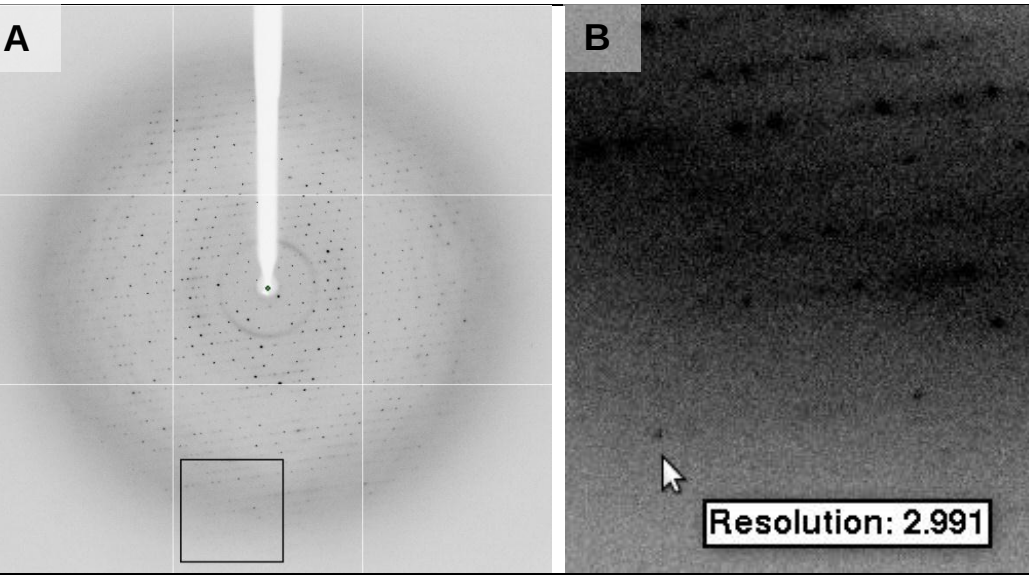


Figure 3.12 A diffraction image of a WT M15-BV crystal. A) Overview and B) Close-up view of a section enclosed by a square.

3.5 Surface entropy reduction strategy

3.5.1 Background

In the previous sections, attempts to solve the structure of Agp1 in various truncated forms have been described. Those efforts have resulted in partially improving the preliminary structure of Agp1, which has been derived from a diffraction dataset described in a previous publication and improved data obtained thereafter (Scheerer et al., 2006). As far as the maximum resolution of diffraction datasets is concerned, 2.77 Å has been the best. Since in the various attempts to crsytallise WT M15-BV only a single crystal form was obtained for which serious problems with structure refinement were observed consistently as described in the previous section, it was deemed necessary to adopt a new strategy to obtain crystals of a different form and a better diffraction quality to enable structure determination.

3.5.2 Current situation surrounding the field of protein crystallography

Since the advent of protein crystallography, solving the structures of proteins of interest continues to be an important aspect of biology in the post-genomic era for detailed insights of the mechanistic function that can be deduced from the coordinates. Despite the increasing wealth of available techniques and recent development of high-throughput methods, obtaining crystals of sufficient quality for diffraction analysis remains to be a big obstacle in protein crystallography today. This is partially due to the fact that the "low-hanging fruits", i.e. more easily crystallised proteins, have been more thoroughly explored. Statistics from structural genomic studies estimate that the likelihood of a given protein entering the pipeline to crystallise is in the range of 30-40%. The probability for the structure to be solved from those crystals is further limited, and ultimately only 10-20% of the proteins entering the pipeline will be successfully structurally determined, thus leaving the majority of the difficult proteins behind (p313, Bergfors, 2009). This is consistent with the fact that the only way to sustain the output from modern structural genomic projects, facilitated by high-throughput technologies, is to input more candidate proteins into the pipeline (Ericsson et al., 2006). It will therefore become increasingly important in the future to directly address this issue that the majority of proteins are not readily crystallisable.

3.5.3 Strategies for increasing the propensity of a protein to crystallise

Many different approaches have been developed for crystallising the protein of interest. The success of protein crystallisation indeed depends on finding the optimal crystallisation condition, but more importantly, the protein itself is the fundamental variable that determines the outcome of the screening (Czepas et al., 2004, Longenecker et al., 2001). Strategies used in modern protein crystallography to modify the target protein include truncation of the construct (Listwan et al., 2009), limited proteolysis (Dong et al., 2007), mutation (Derewenda, 2004b), addition of large tags (Smyth et al., 2003), co-crystallisation with cofactors (Benvenuti and Mangani, 2007), and surface methylation (Walter et al., 2006). To some extent, some of the strategies have already been incorporated in the present study. All attempts to crystallise Agp1 described herein were performed with holoproteins of truncated constructs. The current lack of any crystal structure of apo-phytochrome may be a testament for the necessity of the ligand to stabilise the structure to make the protein amenable to crystallisation. In the present study, apoproteins of mutant M15 proteins were screened but did not give a lead condition (Section 6.1).With regards to the type of chromophore as an optimisation parameter, Agp1 crystals with 15Za have been shown to exhibit better diffraction quality than the holoprotein with the natural chromophore, BV (Scheerer et al., 2010, Scheerer et al., 2006). The effect is probably due to higher conformational homogeneity of the holoprotein with 15Za than the holoprotein with BV, because 15Za in the holoprotein does not undergo the light induced isomerisation, and the holoprotein constitutively assumes a Pr-like state. But the fact that other phytochrome structures with their natural chromophores were refined satisfactorily is a testament that a natural phytochrome holoproteins are capable of forming crystals of sufficient quality, despite being potentially photosensitive. BV was deemed sufficient, and an alternative optimisation strategy was sought.

3.5.4 Background to mutating the protein for crystallisation

It has been known from earlier days of protein crystallography that some proteins are more amenable to crystallisation whereas others are less so. This phenomenon is exemplified by the differences in crystallisation propensities between the protein homologues from different species, and this has been used to an advantage by early day crystallographers in the form of homolog-screening (Derewenda, 2004a). It is a viable option to change the source organism when the protein proves difficult to crystallise

(p79, Blundell and Johnson, 1976). Though working with a protein from another species may yield positive results, it is a time-consuming approach, however. The purification protocol needs to be re-established, and the biochemically active range of the protein may be different. It may therefore be more efficient to modify the existing target protein to improve the crystallisation propensity without altering the structure and the function of the protein. Homologue screening may not be feasible if the protein of interest is a high-value target of human origin with medical relevance. The vast majority of the protein structures solved and deposited in the databases derives from heterologously expressed material, and this approach enables the resulting protein to be modified easily by introducing the change to the expression construct (Derewenda, 2004b).

An early example of successful crystallisation by protein engineering is the case of the human ferritin H-chain. A Lys86Glu mutation was introduced to the human homologue to emulate a Ca^{2+} mediated crystal contact which was found in the crystal of rat homologue of the protein (Lawson et al., 1991).

Surface methylation of amino groups was successfully employed to solve the structure of chicken myosin subfragment-1 (Rayment et al., 1993) and since then it has become a widely accepted rescue strategy for proteins reluctant to being crystallised (Walter et al., 2006).

In the present study, site-directed mutagenesis was chosen over protein methylation for the following reasons. (i) Even if the methylation reaction is perfectly reproducible and reaches completion, i.e. all the amino groups of lysines and the N-terminal methionine residue are methylated, only two variations of the protein are possible as products together with the non-methylated protein. The mutation approach on the other hand will offer the freedom of location, number, and the type of mutation to be introduced, leading to more variations for testing. (ii) Methylation process is unlikely to proceeds to completion in reality. This will lead to inhomogeneity of the sample, which is unfavourable for crystallisation of macromolecules (Benvenuti and Mangani, 2007). (iii) Once the mutated plasmids are produced, the production and purification protocol of mutant proteins are expected to be the same as the established protocol. Producing mutant protein will thus be more cost and time effective.

3.5.5 Mutation strategy employed thus far for crystallisation of other phytochromes

There are three cases where mutating the protein resulted in better diffracting phytochrome crystals (Auldridge et al., 2012, Mailliet et al., 2011, Wagner et al., 2007).

Wagner and co-workers succeeded in improving the maximum diffraction of the crystals of DrBphP from 2.5 Å resolution (1ZTU) to 1.45 Å resolution (2O9C) (Wagner et al., 2007). The presence of chemically impossible overlap between symmetry related Tyr307 was identified in the 1ZTU structure and was assumed to cause a local disorder detrimental to crystallisation (Fig. 3.13). A Tyr307Ser mutant was designed to reduce the conflict and the mutation was justified on the basis of the distance of the mutated residue from the bilin pocket. Mutation of the equivalent residue of Agp1 was first considered as a potential strategy, and in principle this is plausible considering the similarity between currently available phytochrome structures and the high degree of sequence identity between Agp1 and DrBphP for the PAS-GAF domains (37.6 %). Upon inspection of the electron density map of the preliminary M15-BV structure at Phe297, the equivalent of DrBphP Tyr307, the conformation of the phenyl side chains were not identical to that of DrBphP Tyr307, but in fact the side chains of two symmetry related molecules were pointing away from each other. This observation must be taken with care as the calculated electron density of the preliminary structure of M15 is not entirely reliable especially for the side chains. When the side chain conformations for other phytochrome crystal structures in biologically relevant parallel homodimers were studied, the closest distances between the pairs of atoms of the residues were at distances not considered to cause a steric clash as in the case of 1ZTU. Hence it was hypothesised that the steric clashes caused by residues corresponding to Tyr307 in DrBphP phytochrome is not a universal characteristic for all phytochromes, therefore the applicability of this strategy to Agp1 is questionable.

Approaches described by Mailliet et al., 2010 and Auldridge et al., 2012 to mutate the residue(s) necessary for photoconversion in Cph1 and DrBphP, respectively, have led to better diffracting crystals. Mutations to render phytochromes photo-inactive may increase the conformational homogeneity of the sample, but since the ultimate aim of the study is to carry out structural comparison of Agp1 in the Pr- and the Pfr-forms, the mutation to alter the conformation of the functionally important residues was deferred.

Papiz and co-workers successfully crystallised the CBD of RpBphP2 (Bellini and Papiz, 2012a). The WT RpBphP2 CBD was not amenable to being crystallised. Crystallogenesis of RpBphP2 was achieved by introducing the residues involved in crystal contacts in the RpBphP3 structure (Yang et al., 2007). Mutation resulted in successful crystallisation of RpBphP2, albeit with a different crystal form from that of RpBphP3. At the time of planning mutagenesis of Agp1, two PCM structures were available (Cph1, Essen et al., 2008; PaBphP, Yang et al., 2008), but applicability of the approach of Papiz and co-workers may not be as plausible for Agp1 as for RpBphP2. Sequence identities between Agp1, Cph1, and PaBphP (roughly 33%, Table 1.4) is not as high as between RpBphP2 and RpBphP3 (52%). Furthermore, structural differences are greater among phytochrome PCM (PAS, GAF, and PHY domains) structures than CBD (PAS and GAF domains) structures. There are considerable differences in the positions of the PHY domain relative to the PAS-GAF domains in Cph1, PaBphP, and preliminary Agp1 structures (Fig. 3.29), therefore introducing residues participating in crystal contacts in other phytochromes to Agp1 may not result in a successful crystallogenesis.

Since the aim of this study is to gain functional insights from phytochrome structures, attention must be paid to avoid altering the protein function where possible, therefore a more generally applicable approach that does not inadvertently affect the protein's function was sought instead for designing Agp1 mutants.

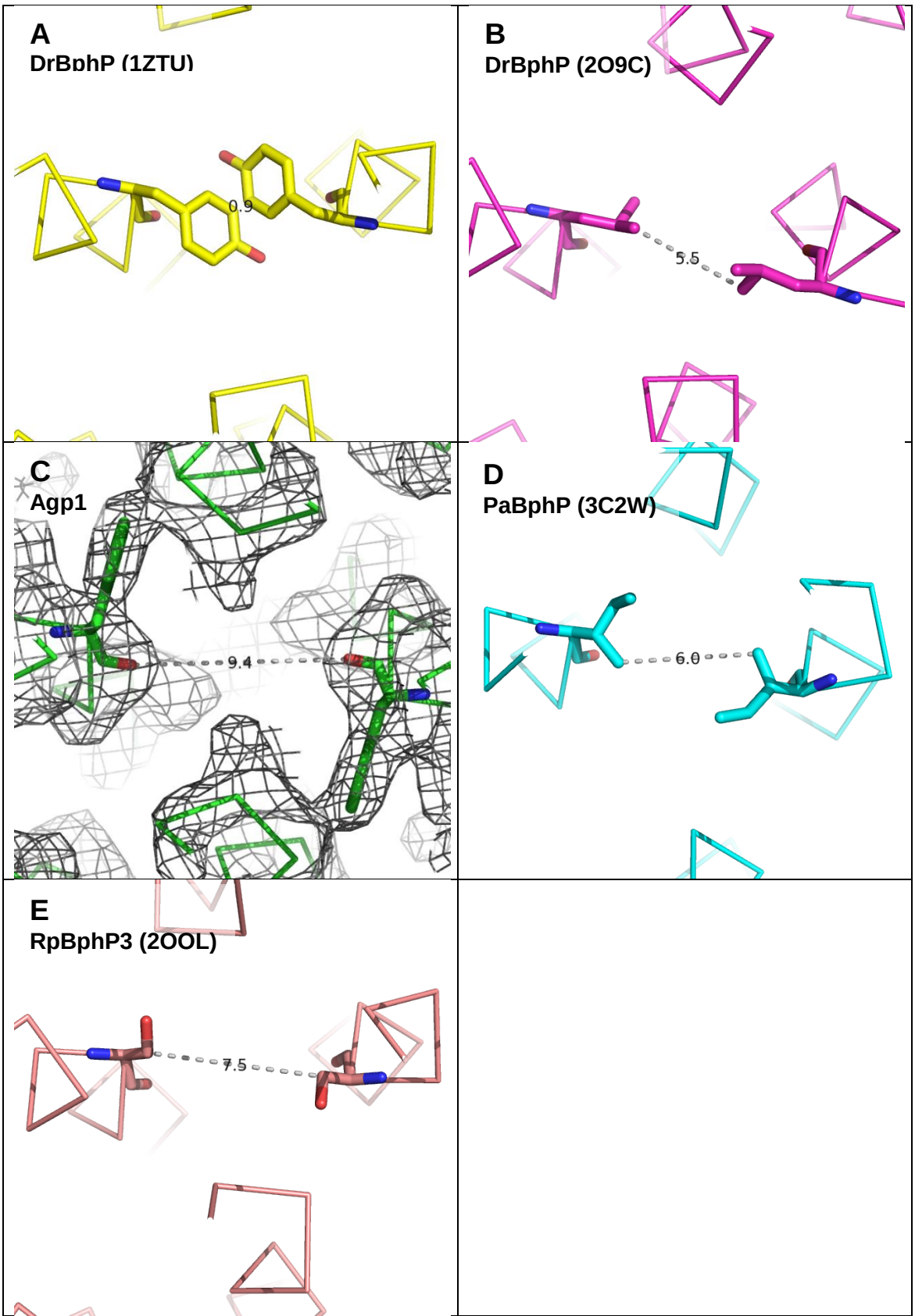


Figure 3.13 Dimer interfaces of DrBphP at Tyr307 and equivalent in other phytochromes. The closest distances between pairs of residues are shown by dotted lines. A) DrBphP (1ZTU), Tyr307, 0.9 Å. This is close proximity between two symmetry related monomers is a physically impossible clash, and the region is disordered in the crystal. B) DrBphP Y307S (2O9C), Ser307, 5.5 Å. C) Agp1 preliminary structure, Phe297, 9.4 Å. D) PaBphP (3C2W), Ile294, 6.0 Å. E) RpBphP3 (2OOL), Leu316, 7.5 Å. The electron density of Agp1 is contoured at 1σ.

3.5.6 Other strategies for Agp1 mutagenesis

It has been shown in the past that it is possible to engineer existing crystal contact sites to improve the crystallisation propensity of proteins. In a systematic study in which all twenty-one mutations at thirteen different crystal packing residues yielded isomorphous crystals under the same condition as the WT, two mutants exhibited exceptionally improved resolution (Mizutani et al., 2008), but the authors conclude that rational prediction for effective crystal engineering is difficult. Another example is a case of increased crystal volumes as a result of mutations which reduced the side chain entropy, and changing the mode of a crystal contact from a solvent mediated one to direct side chain contacts has also been reported (Meher et al., 2009).

Rational engineering of existing crystal contact sites requires the structural information. At the time of planning the mutagenesis strategy, the Agp1 M15-BV preliminary structure was assigned with the amino acid sequence for the PAS and GAF domains while the PHY domain was modelled as a poly-alanine chain (Scheerer et al., 2010). The allocation of correct amino acids in the PHY domain may be deduced by homology modelling, but for the task of predicting the effect of crystal contact engineering, the reliability of the preliminary structure and homology models were deemed insufficient.

In turn a blind approach was adopted instead, using minimal prior knowledge specific to Agp1 or other phytochromes. Such approach may result in obtaining novel crystal forms, and to this end two strategies for mutagenesis were considered.

A strategy to mutate lysine into arginine or glutamine was originally proposed by Dasgupta and co-workers (Dasgupta et al., 1997). In a survey of the contact sites at oligomer interfaces and crystal interfaces from 58 and 223 proteins, respectively, it was suggested that substitution of lysine, the least favoured residue at crystal interfaces, with arginine or glutamine, the two most favoured residues, may promote the formation of crystal contact sites, therefore resulting in crystallisation of a protein.

The “Surface entropy reduction” strategy, originally devised by Derewenda and co-workers, aims to improve the crystallisation propensity of proteins by decreasing the conformational entropy during crystallisation (Derewenda and Vekilov, 2006). The crystallisation process involves transfer of protein molecules from the solution to the crystal at constant temperature and the process is governed by Gibb’s free energy. The

Gibbs-Helmholtz equation at a constant temperature is expressed in the equation below, and Gibb's free energy consists of two components, enthalpy and entropy.

$$\Delta G^{\circ}_{\text{cryst}} = \Delta H^{\circ}_{\text{cryst}} - T\Delta S^{\circ}_{\text{cryst}}$$

Where G°_{cryst} is Gibb's free energy, H°_{cryst} is enthalpy, S°_{cryst} is entropy and Δ indicates the difference in any of these for the system before and after the crystallisation. The crystallisation process is thermodynamically favoured when the change in Gibb's free energy is negative as a result of contribution from both enthalpy and entropy. Given that only moderately negative changes in Gibb's free energy, usually in the range of -10 to -100 kJ mol^{-1}, are associated with protein crystallisation (Vekilov, 2003), and changes in enthalpy are often either moderate (-70 kJ mol^{-1} for lysozyme (Schall et al., 1996)) or negligible (~0 kJ mol^{-1} for ferritin, apoferritin, and lumazine synthase (Gliko et al., 2005, Petsev et al., 2001, Yau et al., 2000)), the crystallisation process must be predominantly entropy driven. The entropy component in the above equation consists of entropic contributions from both protein and water molecules forming a layer of structured water surrounding the protein, therefore the equation may be further extended to the form shown below.

$$\Delta G^{\circ}_{\text{cryst}} = \Delta H^{\circ}_{\text{cryst}} - T(\Delta S^{\circ}_{\text{protein}} + \Delta S^{\circ}_{\text{solvent}})_{\text{cryst}}$$

Some amino acids contribute to an increase in the Gibb's free energy during the crystallisation process. This is because when the residues with many conformations (high entropy) in solution are stabilised as a result of the crystallisation, the drop in conformational entropy (TS_{protein}, a negative value) increases the Gibb's free energy due to the negative operation on the entropy in the equation given above. Replacing residues which decrease the conformational entropy during the crystallisation ("low entropy residues") with alanine with no conformational entropy will in theory make the crystallisation process more thermodynamically favourable (Table 3.5).

Residue	TS_{coil} (kcal / mol)
Ala	0.000
Gly	0.000
Pro	0.000
Val	-0.541
Phe	-0.544
Asn	-1.708
Glu	-1.763
Lys	-1.973
Gln	-2.107
Arg	-2.120

Table 3.5 The conformational entropy of selected amino acids multiplied with temperature (300 K). Five amino acids with the lowest and highest conformational entropy-dependent change in energy upon crystallisation in coil regions (TS_{coil}) are shown. Adapted from Avbelj and Fele, (1998).

Lysine and Glutamate are the primary mutable residues in this strategy. Lysine occupies 12 – 15% of protein surfaces, yet only 5.4% of oligomer interfaces. This indicates that the presence of lysine is disfavoured at protein-protein interaction sites. Like lysine, the presence of glutamate is deferred at interfaces between proteins. Glutamate occupies 10.3% of protein surfaces, yet it constitutes only 4.1% of oligomer interfaces. It must also be noted that glutamate is present predominantly on the surface of the protein and that only 12% of Glu residues within the polypeptide are buried. Mutating glutamine with alanine will provide entropic benefits without decreasing the protein solubility significantly, as the change in charge is smaller than Glu→Ala or Lys→Ala mutations. Glutamine is less suited as a SER target residue, however, as a higher proportion of glutamine residues are present in the protein interior (17%) compared to glutamate (12%), and glutamine constitutes a smaller proportion in proteins (3.7%) which makes glutamine a less suitable target for a blind approach (Derewenda and Vekilov, 2006). Arginine is not regarded as a target mutable residue, because oligomer interfaces actually favour arginine so that the strategy by Dasgupta recommends mutating lysine into arginine and glutamine. To the best of the author's knowledge, no case of successful SER approach targeting asparagine has been reported, despite its fifth lowest conformational entropy in coil (Table 3.5).

Therefore the primary targets of mutable residues are lysine and glutamate because of their high probability to occur at the surface of the protein but not at oligomer interfaces, and mutating them into alanine will make crystallisation more thermodynamically favourable. A successful case of SER approach to target aspartate residues has also been reported (Cierpicki et al., 2006), but aspartate is not the primary target due to its frequent presence at protein-protein interaction sites.

Numerous successful cases have been reported for both strategies by Dasgupta et al. and by Derewenda et al. To the best of the author's knowledge, there is only one study in which two strategies were directly compared for their impact on crystallisation (Czepas et al., 2004). Therein it was concluded that Dasgupta's approach improved the crystallisation propensity of the protein, but the effect was not as dramatic as SER approach.

In the present study, the SER approach was adopted as a method of improving the quality of Agp1 crystals. SER mutations were only carried out on Lys and/or Glu residues occurring consecutively without a gap, because the beneficial effect to crystallisation is known to be more pronounced than when the mutations were introduced to a single residue (Cooper et al., 2007).

3.6 Single cluster surface entropy reduction mutants

3.6.1 Generation of single cluster surface entropy reduction mutants

3.6.1.1 Target site determination

Initial candidates of mutants according to the SER strategy were designed simply by identifying two or more consecutive lysine and/or glutamate residues, and the maximum gap allowed was a single residue. Based on this simply criterion, nine initial candidates were identified (Fig. 3.14). Hereafter the SER candidate clusters will be referred to by the order in the sequence from the N-terminus, i.e. SER2 refers to E86A/E87A. SER1 (E7A/K8A) was immediately removed from the list of candidates, as plasmid pNM19a which lacks the coding sequence corresponding to the first nine N-terminal residues (Section 3.4) was the base to generate the mutated constructs. Three final candidates were selected from the initial candidates according to the following additional criteria.

Firstly, the mutation must only be introduced in the loop region on the surface of the protein. The SER approach is more successful when the mutation is introduced to loops rather than α-helices (Goldschmidt et al., 2007). Residues that do not occur on the protein surface will not influence the entropy upon crystallisation, except for an unforeseen conformational change that is associated with the mutation of buried residue, which is hard to predict. The secondary structure of the regions containing the candidate residues was checked by inspecting the preliminary structure of Agp1. For regions where the structural information was missing and/or the primary structure was not assigned (poly-alanine model region), a homology model of Agp1 based on the crystal structures of Cph1 (2VEA) and DrBphP (2O9C) was used. Only SER2 (E86A/E87A), SER3 (E96A/K97A/K98A), and SER6 (E336A/K337A) were found likely to occur in loop regions, and the degree of surface exposure of these residues in the candidate sites was confirmed with Areaimol (Saff and Kuijlaars, 1997).

Secondly, the residues to be mutated must not be conserved amongst other phytochromes to minimise the potential functional effect the mutation may introduce. To this end, a sequence alignment of Agp1 with other phytochromes was utilised.

Finally, the candidate residues were checked for any report of functional alteration upon mutation of the corresponding residues in other phytochromes. To this end, as many mutational studies on phytochromes as possible were compiled, and based on the sequence alignment, the residues corresponding to the candidate SER residues were

checked for previously reported mutational studies. The candidate mutation sites and the corresponding residues in other phytochromes are shown in Table 3.6. To the best of the author's knowledge, no mutation has been carried out on the residues corresponding to the SER targets to date. However, the lack of evidence for the functional importance of the target residues is not the evidence that mutating those residues will not affect the protein function.

Alongside the manual selection of the mutable residues, the sequence of Agp1-M15s was submitted to the SERp server (http://services.mbi.ucla.edu/SER/) for automated identification of mutable residues (Goldschmidt et al., 2007). Thereby three mutable clusters were identified, and these clusters were identical to SER3, SER6, and SER2, in order of decreasing priority. The identification of the same clusters as mutable residues by both manual selection and the automated search is not surprising considering that the criteria used in the processes were essentially the same. Consequently, three candidate clusters, SER2 (E86A/E87A), SER3 (E96A/K97A/K98A), and SER6 (E336A/K337A) were selected for carrying out SER mutations to Agp1 (Fig. 3.15).

3.6.1.2 Generation of the plasmid constructs

Three single cluster SER mutations were introduced to the pNM19a plasmid according to the QuikChange protocol (Section 2.2.2.9) and the procedure is described in Appendix 3.

MQRERLEKVMSSHTPKLDSCGAEPIHIPGAIQEHGALLVLSAREFSVVQASDNLANYIGVDLPIGAVATE
ANLPFISVLSAWYSGEESNFRYAWAEKKLDVSAHRSGTLVILEVEKAGVGESAEKLMGELTSLAKYLNSA
PSLEDALFRTAQLVSSISGHDRTLIYDFGLDWSGHVVAEAGSGALPSYLGLRFPAGDIPPQARQLYTINR
LRMIPDVDYKPVPIRPEVNAETGAVLDMSFSQLRSVSPVHLEYMRNMGTAASMSVSIVVNGALWGLIACH
HATPHSVSLAVREACDFAAQLLSMRIAMEQSSQDASRRVELGHIQARLLKGMAAAEKWVDGLLGGEGERE
DLLKQVGADGAALVLGDDYELVGNTPSREQVEELILWLGEREIADVFATDNLAGNYPTAAAYASVASGII
AMRVSELHGSWLIWFRPEVIKTVRWGGDPHKTVQESGRIHPRKSFEIWKEQLRNTSFPWSEPELAAAREL
RGAIIGIVLRKTEE

Figure 3.14 Clusters of mutable residues identified within Agp1-M15 sequence. Two or more consecutive Lys or Glu residues with maximum gap of one residue are highlighted. Initial candidate SER mutation clusters were, SER1, E7A/K8A; SER2, E86A/E87A; SER3, E96A/K97A/K98A; SER4, E113A/E115A/K116A; SER5, E124A/K125A; SER6, E336A/K337K; SER7, E382A/E383A; SER8, K469A/E470A; SER9, K501A/E503A/E504A.

Agp1		PaBphP		RpBphP2		RpBphP3	
Amino acid	Residue number	Amino acid	Residue number	Amino acid	Residue number	Amino acid	Residue number
E	86	W	82	I	91	I	105
E	87	S	83	A	92	R	106
E	96	E	92	D	100	E	115
K	97	H	93	A	101	R	116
K	98		N/A	G	102	A	117
E	336	D	333	H	341	Q	355
K	337	D	334	D	342	G	356

DrBphP		Cph1		*A. thaliana* PhyA		*A. thaliana* PhyB	
Amino acid	Residue number	Amino acid	Residue number	Amino acid	Residue number	Amino acid	Residue number
A	96	S	92	I	156	V	190
L	97	K	93	L	157	W	191
P	106	D	102	P	166	P	200
A	107	F	103	F	167	F	201
A	108	V	104	Y	168	Y	202
L	346	A	346		N/A		N/A
S	347	D	347		N/A		N/A

Table 3.6 SER-target residues in Agp1 and corresponding residues in other phytochromes. No mutation, whether naturally occurring or introduced artificially, to the residues corresponding to the SER targets in Agp1 has not been reported in any phytochromes.

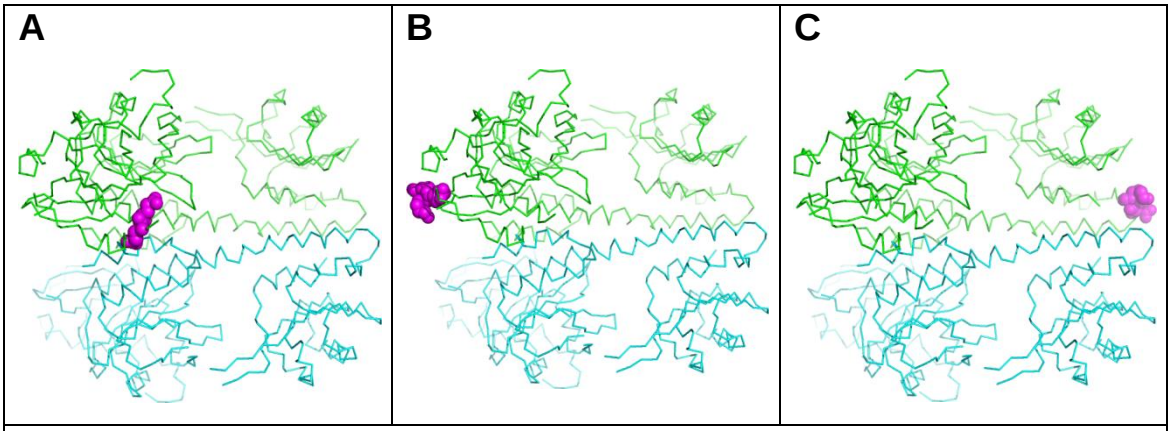


Figure 3.15 Location of the SER target clusters. A preliminary structure of M15 is displayed as a dimer in C_α representation, with one monomer coloured green and the other coloured cyan. The target clusters are displayed as violet spheres. Location of the A) SER2 (E86A/E87A), B) SER3 (E96A/K97A/K98A), and C) SER6 (E336A/K337A) clusters relative to the global protein structure.

3.6.2 Expression and purification of single cluster SER-BV protein

Apoproteins of SER2, SER3, and SER6 mutants were overexpressed with 1.5 L of RB-Amp media. Apart from the significantly reduced yield of apoprotein for SER3 and SER6, no difference in the characteristics of the protein was observed during affinity purifications, so that the mutant apoproteins were successfully purified following the previous protocol (Section 3.4.1). Samples during the expression and affinity purification of the mutants were visualised by SDS-PAGE gels (Fig. 3.16). Presence of SER2 protein in the flow-through during sample application and washing stage may indicate inefficient binding of the protein to the Ni^{2+} charged resin, but this was attributed to extremely high level of soluble protein which saturated the binding capacity of the column, and this is consistent with relatively high level of SER2 yield compared to other SER mutants.

For each SER mutant the yield after affinity purification was different from that of the WT protein (data not shown), suggesting that the solubility of M15 proteins changed following SER mutations. The effect of SER mutations on the solubility of the proteins is discussed in more detail in Section 5.2.

Affinity purified SER mutant apoproteins were subjected to the same protocol as the WT M15s protein with regards to the assembly with BV and the following purification with SEC (Section 3.4.1). Purified holoprotein was concentrated to 25 mg/ml for screening.

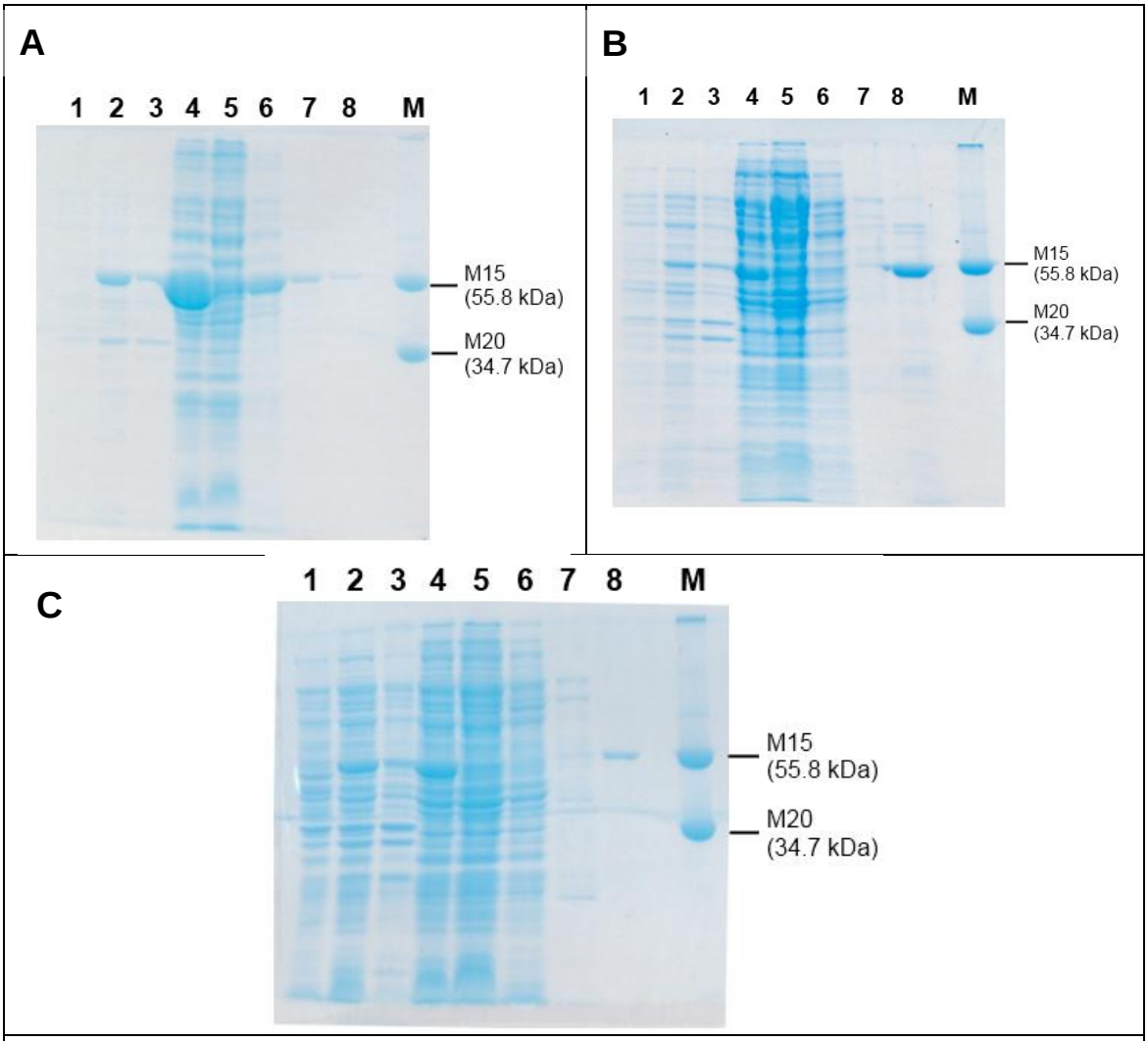


Figure 3.16 SDS-PAGE gels (10%) showing the expression and affinity purification of single cluster SER mutants. A) SER2. B) SER3. C) SER6. In all cases, the purification stages in the individual lanes are as follows. Lane 1, cell pellet before induction; Lane 2, cell pellet after induction; Lane 3, insoluble fraction after lysis; Lane 4, soluble fraction after lysis; Lane 5, column flowthrough when the cleared soluble fraction was applied onto the affinity column; Lane 6, first half of the flowthrough from the affinity column during the wash, Lane 7, second half of the flowthrough from the affinity column during wash; Lane 8, purified protein eluted from the affinity column; M, SN marker. Lane 8 of the first gel appear faint due to low quantity of protein being loaded into the well.

3.6.3 Crystallisation screening with purified single cluster SER-BV protein

Following the successful production and purification of the SER2, SER3, and SER6 mutants, they were characterised by multitudes of biochemical approaches prior to crystallisation trials (Section 5.3).

De novo screening was carried out on BV adducts of the SER mutants (SER-BV) in order to maximise the chance of obtaining new crystal forms. Screening was carried out with JCSG+, PACT-premier, and PGA kits (Molecular Dimensions) along with the WT M15s-BV protein as control. The protocol used was identical to that for the WT M15s-BV protein (Section 3.4.2).

Crystallisation trials with the SER-BVs initially resulted in re-discovering the crystallisation conditions already identified for the WT protein. An exception is SER3-BV, which was later found to crystallise in a novel crystal form. Since this novel form, however, was first obtained with another SER mutant, SER36-BV, crystallisation of SER3-BV will not be discussed in detail here.

3.6.4 Analysis of single cluster SER-BV crystals

Many crystallisation conditions have been identified for the BV adducts of SER2, SER3, and SER6, but most of the conditions consisted of PEG of intermediate molecular weights (3350-8000) with salts containing mono/di-valent cation in a mildly alkaline range of pH 7.5 – 8.5, which are essentially variations of the crystallisation condition for the WT M15-BV holoprotein (Section 3.4.2). Crystals of the single cluster SER mutants with a morphology indistinguishable from the WT holoprotein crystals appeared under the optimised condition for the WT holoprotein. SER2-BV crystals exhibited comparatively favourable morphology featuring high volume per length, and the crystals were subjected to diffraction analysis at the ESRF synchrotron.

A dataset termed 5-3 with relatively good diffraction quality was collected from a SER2-BV crystal at ID14-1 at the ESRF synchrotron (Fig. 3.17). The dataset termed 5-3 was indexed, scaled, and processed with the same protocol as used for the dataset from the WT holoprotein crystals (Section 3.4.3). The scaling statistics and L-test for twinning are as shown in Table 3.7 and Fig. 3.18, respectively, and they indicate the quality to be not as good as that of dataset 11-2 of WT M15-BV crystal (Section 3.4.3). Given that SER2-BV crystals exhibit a similar morphology to the WT crystals, it is not surprising that the unit cell parameters and the apparent space group are essentially the

same. Structural refinement was attempted following the phase determination with molecular replacement using the space group I422, but a case of "pathological" crystal was evident as the structural refinement proved to be difficult similarly to the case of WT (Data not shown). Given the persistence of crystallographic difficulties in SER2-BV crystals, no further attempts were made to solve the structure with single cluster mutant of the protein.

Screening of single cluster SER mutant proteins led to identifying similar crystallisation conditions to that of the known conditions for the WT M15s. Furthermore, it has been shown that the crystal form of SER2 crystal to be the same as the WT crystal, and the same is assumed to apply to other SER mutant crystals judging by the crystal morphology and the condition in which crystals grow.

It has been confirmed from crystallisation condition, crystal morphology, and more directly by data processing that SER2-BV holoprotein crystal is in the same "pathological" form to that of the WT holoprotein, thereby resisting satisfactory structural refinement. It is also likely for SER3-BV and SER6-BV adducts to be in the same crystal form, and for this reason, more extensive changes were judged to be necessary for inducing crystallisation in a novel form.

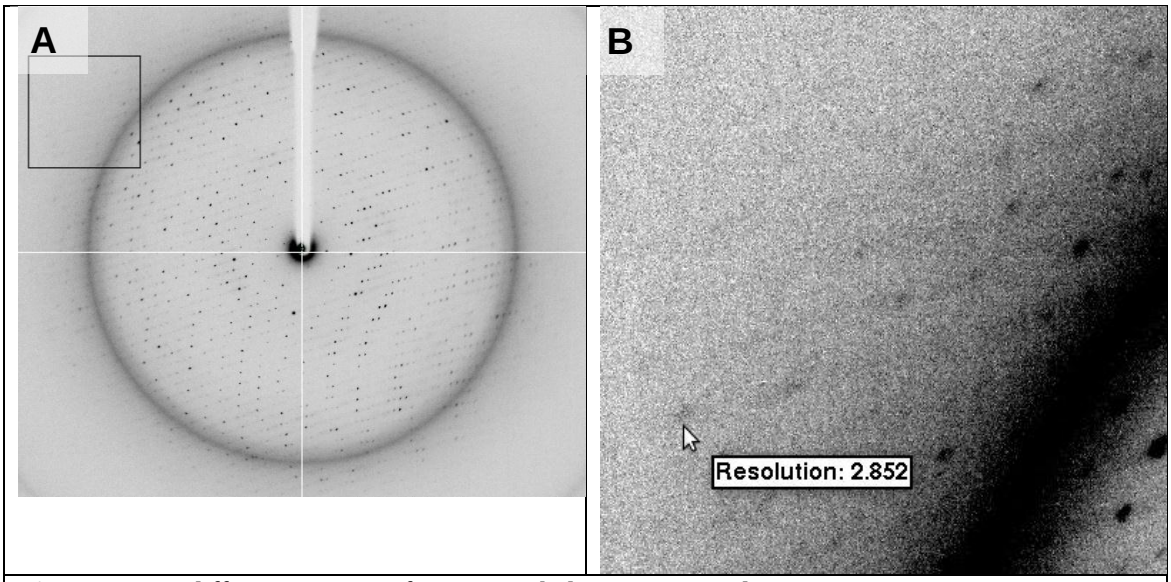


Figure 3.17 A diffraction image of SER2-BV holoprotein crystal.

Protein (Dataset)	SER2-BV (5-3)
Beamline/wavelength [Å)	ID14-1 / 0.9334
Space group	I422
Unit cell dimensions	$a = b$ = 171.06 Å, c = 78.52 Å, α = β = γ = 90 °
Resolution range (outer shell) [Å]	54.79 – 2.76 (2.91 – 2.76)
No. of observations	64254
No. of unique reflections	13305
Multiplicity	4.8
Completeness overall (outer shell) [%]	88.3 (80.9)
<I/σ(I)> overall (outer shell)	7.9 (2.2)
Rmerge overall (outer shell)	0.107 (0. 677)
Rmeas overall (outer shell)	0.120 (0.769)
Rpim overall (outer shell)	0.052 (0.352)
Mosaicity (°)	1.91
Estimated B-value (Wilson plot) [Å^2]	58.4

Table 3.7 Crystal data of dataset 5-3 of a SER2-BV crystal.

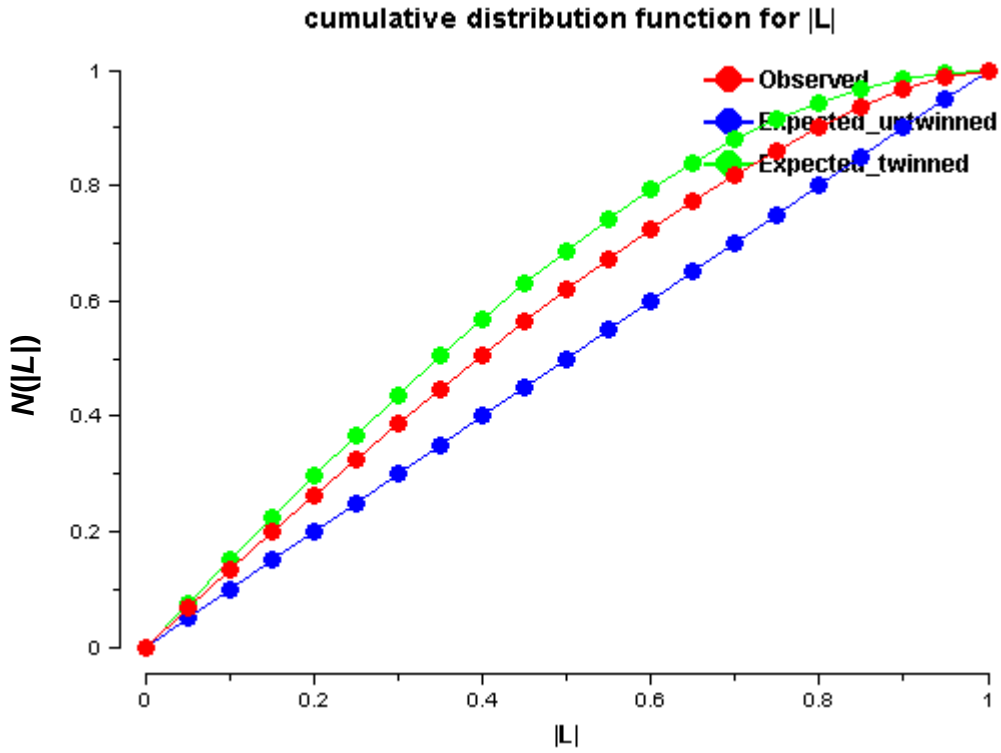


Figure 3.18 L-test for twinning of dataset 5-3 of SER2-BV crystal. Twin fraction of the dataset 5-3 from a SER2-BV crystal is higher than the dataset 11-2 from a WT crystal (Fig. 3.11).

3.7 Double cluster surface entropy reduction mutants

Screening single cluster SER mutants resulted in identification of similar crystallisation condition to the WT Agp1 M15-BV protein. The crystal form of the WT was preserved in all single cluster SER mutant crystals as inferred by morphologies, as well as by analysing the diffraction images from SER2-BV crystals, thus the crystals of the single cluster SER mutants are expected to be in the "pathological" crystal form, except for SER3-BV crystal which was later found to crystallise in an $(NH_4)_2SO_4$-based condition with a novel form (Section 5.4). Introduction of single clutter SER mutation did not result in drastic alteration of crystallisation propensity of the protein, and furthermore it was not possible to improve the quality of the crystals by crystallographic techniques, such as re-crystallisation (Wagner et al., 2005), streak/micro seeding, or by altering the crystallisation configuration to sitting-drop. Hence in an effort to obtain crystals of BV adducts of Agp1 PCM in alternative form, simultaneous introduction of two SER mutations was attempted, as it was hypothesised that additional SER mutation may induce new crystal form by further shifting the thermodynamically stable intermolecular interaction *in crystallo*.

3.7.1 Generation of double cluster surface entropy reduction mutants

Two double cluster SER mutants, SER26 and SER36 were arbitrarily chosen for production. To carry out mutagenesis, pNM19a plasmid bearing SER6 mutation was chosen as the template. SER6 plasmid was subjected to mutagenesis with SER2 F and SER2 R primers to produce SER26, and with SER3 F and SER3 R to produce SER36 mutant according the QuikChange protocol in Section 2.2.2.9. Plasmids were purified from successfully transformed colonies, and successful mutagenesis was verified as described in Appendix 4.

3.7.2 Expression and purification of double cluster SER-BV protein

Double cluster SER mutants were produced and affinity purified following the identical protocol to that of the WT and single cluster SER mutant proteins (Fig. 3.19, Section 3.4.1 and 3.6.2). Introduction of SER2 or 3 mutations decreased the solubility of the SER6 protein like when SER2 and 3 single cluster mutations were introduced to the WT protein (Section 5.2). Apoproteins were assembled with BV following the same protocol as for the WT M15s, and subjected to SEC with Superdex 200 column (Section 2.2.3.7.6).

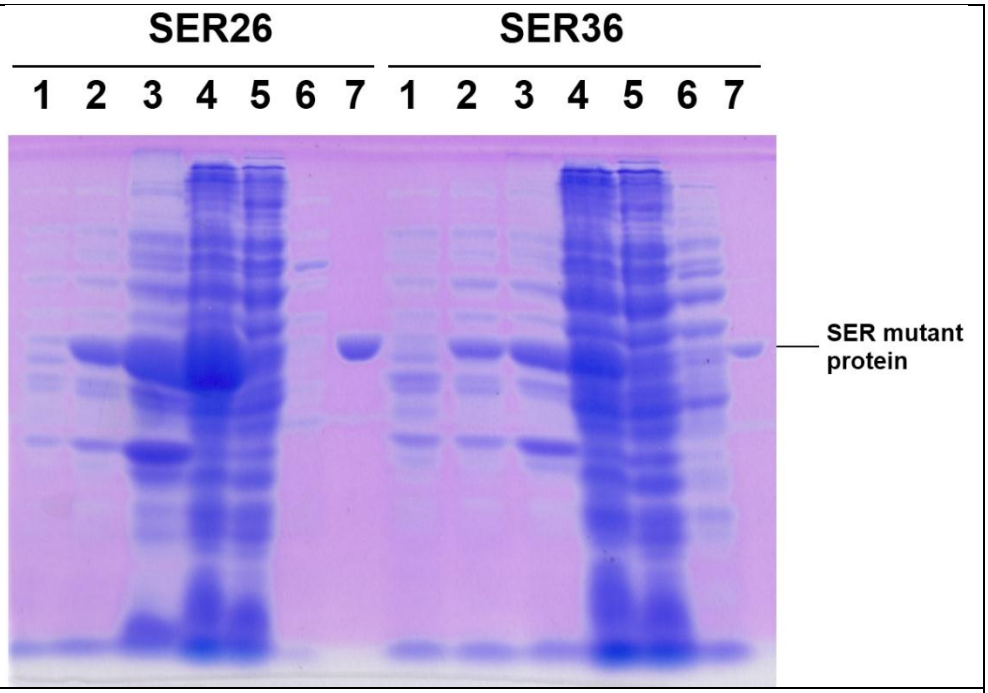


Figure 3.19 SDS-PAGE gels showing the expression and affinity purification of SER26 and SER36 mutants. Various purification stages indicated by the numbers were separated by SDS-PAGE (10%). Lane 1, Cell pellet before induction; Lane 2, Cell pellet after induction; Lane 3, Insoluble fraction after lysis; Lane 4, Soluble fraction after lysis; Lane 5, Column flowthrough during applying the cleared soluble fraction onto the affinity column; Lane 6, Column flowthrough from washing , Lane 8, Purified protein eluted from the affinity column.

3.7.3 Crystallisation screening with purified double cluster SER-BV proteins

Purified SER26-BV and SER36-BV proteins were used in *de novo* crystallisation screening with JCSG+, PACT-premier, PGA, and Morpheus kits (Molecular Dimensions) according to the standard protocol (Section 2.2.6.1). Lead conditions were identified for both SER26-BV and SER36-BV and optimisation was carried out based on those conditions. The optimised conditions are shown in Table 3.8.

Following the identification of a new crystal form with SER36-BV protein in an (NH4)2SO4-based condition, SER3-BV was also found to crystallise in a similar novel form under the same crystallisation condition (JCSG+ 2.4). No diffraction images of SER3-BV crystals could be collected at a synchrotron, however, due to the time constraints.

Crystals suitable for diffraction analysis were obtained in a condition containing 2 µl each of protein solution at 25 mg/ml and the reservoir solution, and imperfect tetragonal bipyramidal crystals of SER26-BV and SER36-BV crystals with less defined morphology appeared within 2 weeks of setup (Fig. 3.20). Reproducibility of the crystal was high for SER36-BV, but only moderate for SER26-BV crystal.

Introduction of SER mutation clusters resulted in novel crystal forms, where SER26 crystallised in a novel form, and SER3 and SER36 crystallised in another novel form, as judged from the new crystallisation conditions and new crystal morphologies.

Protein	Reservoir solution composition	Temperature (°C)
SER26-BV	0.03 M diethyleneglycol, 0.03 M triethyleneglycol, 0.03 M tetraethyleneglycol, 0.03 M pentaethyleneglycol, 20% Glycerol, 10% PEG8000, 0.05 M MES, 0.05 M Imidazole pH 6.5	25
SER3-BV SER36-BV	1.1-1.2 M $(NH_4)_2SO_4$ 50-200 mM Li_2SO_4 0.1 M Tris-HCl pH 8.5	14

Table 3.8 Optimised crystallisation conditions of SER26-BV and SER36-BV protein.

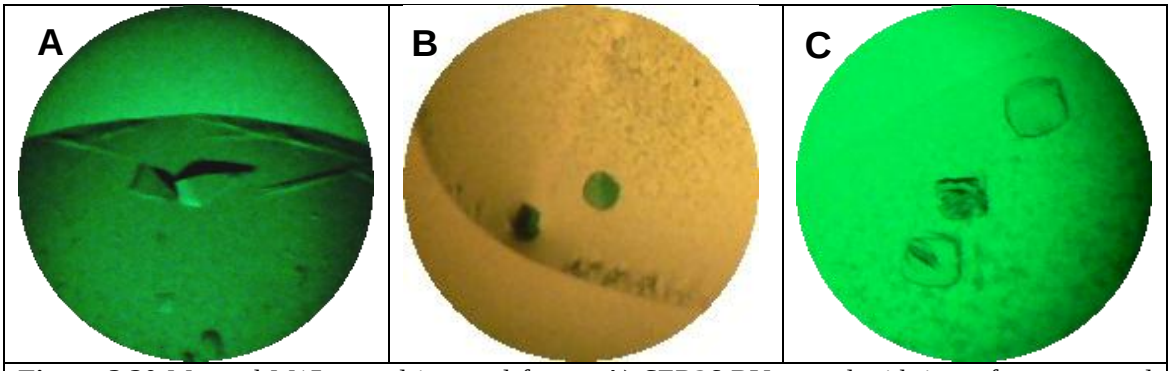


Figure 3.20 Mutated M15 crystal in novel forms. A) SER26-BV crystal with imperfect tetragonal bipyramidal form. SER3-BV B) and SER36-BV C) crystals with relatively undefined morphology.

3.8 Structural determination of SER26-BV

3.8.1 Analysis of double cluster SER-BV crystals

Crystals of SER26-BV and SER36-BV were analysed at ID14-1 and ID14-4 at the ESRF synchrotron (Fig. 3.21). Diffraction images obtained from SER36-BV crystals were of poor quality and no dataset was collected, but diffraction images from SER26-BV were of superior quality and isotropic than that of the WT M15-BV crystal, and the crystal data are summarised in Table 3.9.

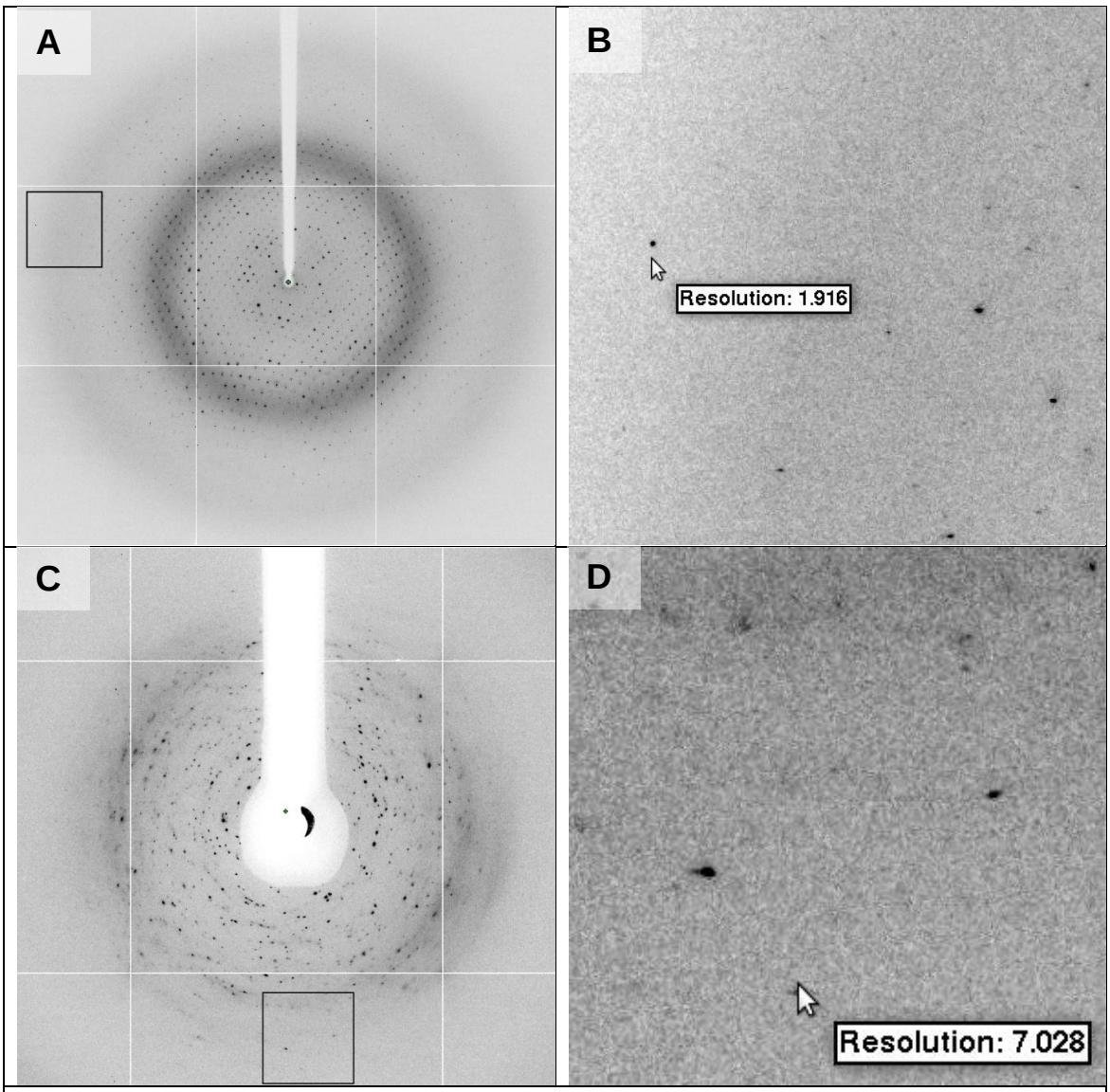


Figure 3.21 Diffraction images from crystals of mutated M15 proteins. A) A diffraction image of SER26-BV, exhibiting improved maximum diffracting resolution, and B) the up-close view of the delimited region. C) A diffraction image of SER36-BV, and D) the up-close view of the delimited region.

Protein (Dataset)	SER26-BV (4-1)
Beamline / wavelength [Å]	ID14-4 / 1.0718
Space group	$P4_32_12$
Unit cell dimensions [Å, °]	$a = b = 69.35$, $c = 236.85$ $\alpha = \beta = \gamma = 90$
Resolution range (outer shell) [Å]	49.04 – 1.82 (1.92 – 1.82)
No. of observations	714810
No. of unique reflections	52618
Multiplicity	13.6
Completeness overall (outer shell) [%]	99.8 (100.0)
$\langle I/\sigma(I) \rangle$ overall (outer shell)	20.2 (4.2)
Rmerge overall (outer shell)	0.080 (0.602)
Rmeas overall (outer shell)	0.083 (0.626)
Rpim overall (outer shell)	0.023 (0.173)
Mosaicity (°)	0.19
Estimated B-value (Wilson plot) [$Å^2$]	31.9
R/R_{free} (%)	21.90/25.87
Number of water molecules per ASU	239
R.m.s. deviations from ideal geometry	
Bond lengths [Å]	0.026
Bond angles [°]	2.126
Chiral volumes [$Å^3$]	0.151

Table 3.9 Summary of data-collection and refinement statistics.

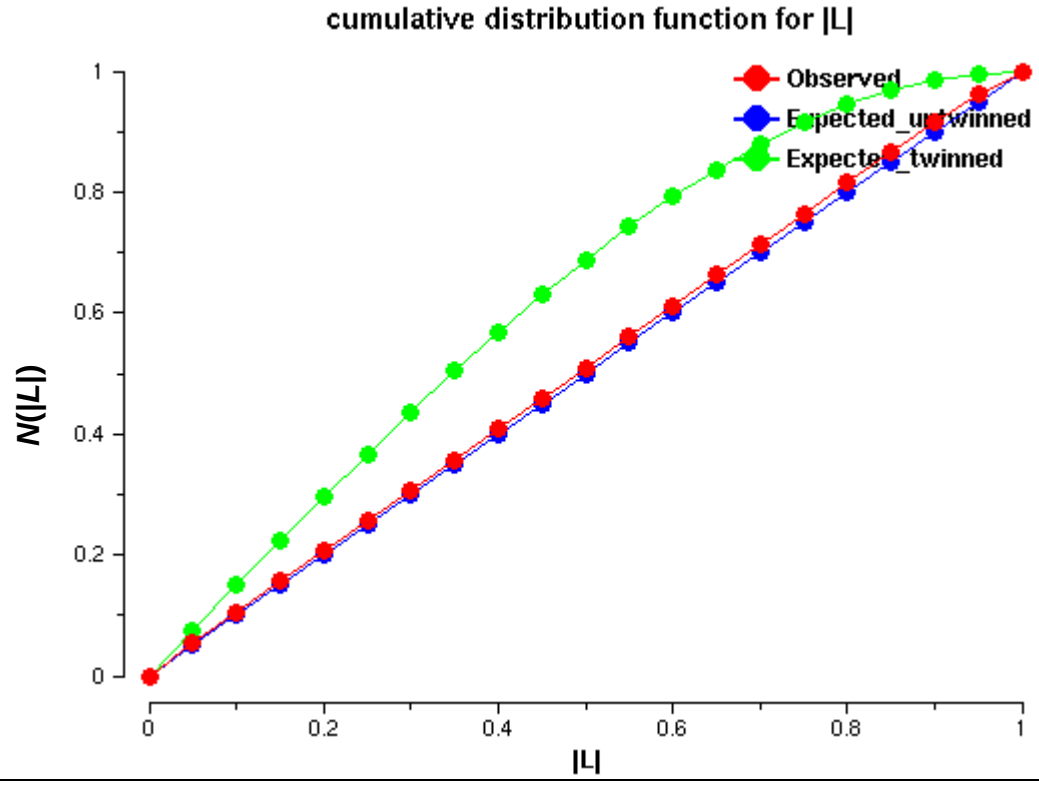


Figure 3.22 L-test for twinning on SER26-BV crystal. The observed values are closer to the values of "Expected untwinned" than WT M15-BV (Fig. 3.11) and SER2-BV (Fig. 3.18).

3.8.2 Data processing and refinement

Phases were determined by molecular replacement with Phaser (McCoy et al., 2007) using the high resolution structure of DrBphP (2O9C) and the PHY domain (331-514) of Cph1 (2VEA). The structure was refined by cycles of automated and manual refinement with Refmac5 (Murshudov et al., 2011) and Coot (Emsley et al., 2010). The library file necessary for defining the ideal geometry of the covalently bound biliverdin was derived as described in Section 2.2.7.2.4. Three possible geometries of the methyl side chain was tested (Section 3.10.4), and the covalent bond distance between sulphur-gamma atom and $C3^2$ atom of BV was defined to be 1.790 Å as in DrBphP structure (2O9C, Wagner et al., 2007).

Refinement of SER26-BV structure resulted in a model consecutively defined for residues Glu 18 to Glu 504 of the Agp1 primary structure. Disordered loop regions were omitted, which were Met 10 – Leu 17, Leu 505, Glu 506, and His 507- His 512. Residues Leu 505 – Glu 506 and His 507 – His 512 correspond to translated XhoI restriction site and the C-terminal hexahistidine tag, respectively.

Refinement statistics are detailed in Table 3.9. The structure was refined to R/R_{free} values of 0.2190/0.2587. The number of water molecules per asymmetric unit is 239, which is small considering the relatively high resolution of the dataset and the molecular weight per asymmetric unit. This is due to a conservative estimation at the current stage of refinement, and more water molecules will be identified in the final refinement cycles.

In addition to the BV molecule in the bilin pocket of the GAF domain, another ligand was found in a cavity between the GAF and the PHY domains. The density of the ligand is neighboured by Asp167, Phe168, Gly169, Leu170, Trp274, and Gln310. The identity of the ligand was judged to be a Tris molecule from the protein buffer based on the molecular weight and the presence of a tetrahedral atom inferred from the electron density (Fig. 3.23). The judgement was supported by the "Ligand identification" module in Phenix which listed Tris as the most likely ligand out of all the reagents present in the crystallisation drop.

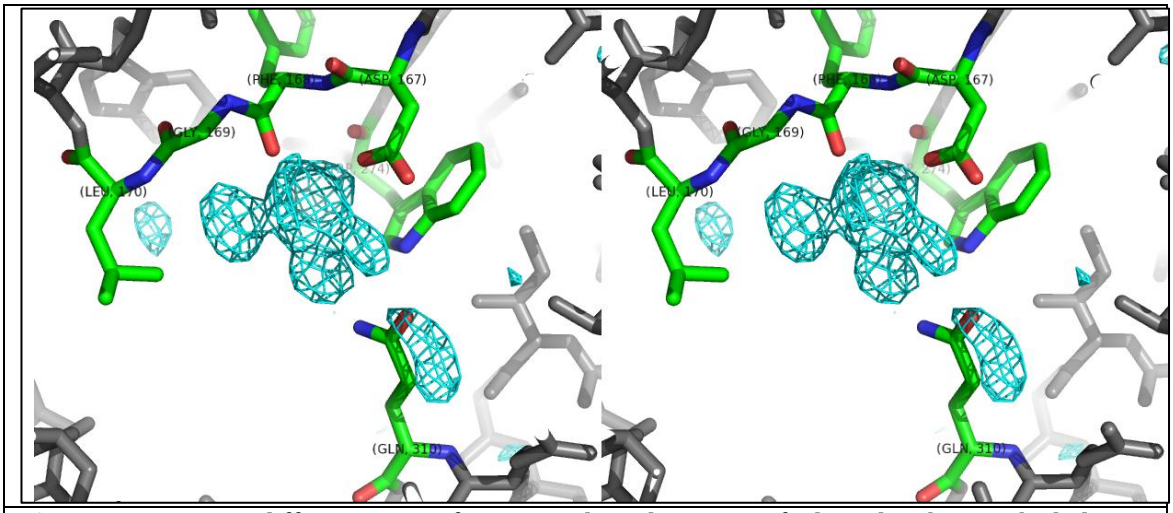


Figure 3.23 Positive difference map of a putative ligand. Density of a ligand with a tetrahedral atom in the vicinity of Asp167, Phe168, Gly169, Leu170, Trp274, and Gln310 is displayed in cyan mesh. Picture is presented in cross-eye stereo. The $F_o - F_c$ map is contoured at 3σ.

3.8.3 Assessment of the extent of radiation damage

Radiation damages at the covalent link between the BV and the conserved Cys residue have been reported in bacteriophytochrome CBD structures (Wagner et al., 2005, Yang et al., 2007). Presence of similar radiation damage in Agp1 was investigated with dataset 4-1. From the original entire dataset that covers 180 ° over 900 images, the first and the last 400 images were independently merged, scaled, and subjected to molecular replacement with the structure of SER26-BV, followed by a round of refinement by Refmac5. 400 images cover 80 ° that contain 99.9% of the unique data with average multiplicity of 2.3. The resulting electron densities from two datasets are shown in Fig. 3.24. Since no clear difference is observed between the two in the region encompassing the covalent bond, no substantial radiation damage is believed to have occurred during the data collection, and justifies merging the entire images in favour of high multiplicity.

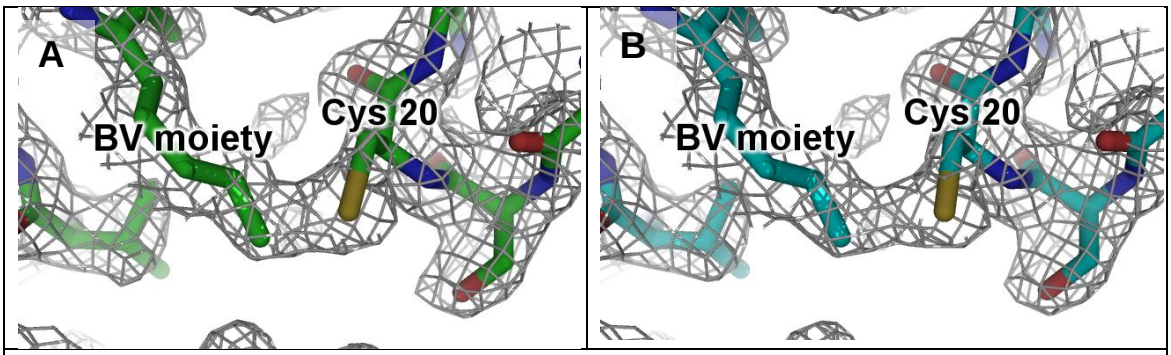


Figure 3.24 Electron density maps from two sections of the dataset processed independently. Electron density calculated from the first 400 images A) and the last 400 images B) are indistinguishable when viewed at the same map contour level at 1σ. The covalent bonds between the BV moiety and Cys 20 have been omitted from the figures.

3.9 An attempt to solve the structure of WT Agp1 with the high resolution mutant structure

Successful structural determination of SER26-BV opened an opportunity to use the reliable mutant structure for the molecular replacement for very accurate phase determination. Thereby it will be possible to completely remove one of the problems which may have been preventing the refinement of the WT structure.

3.9.1 Molecular replacement and refinement

The merged and scaled diffraction dataset 11-2 from a WT M15-BV crystal was subjected to phase determination by molecular replacement using the refined structure of SER26-BV but without the BV. A large scale conformational change was found to be associated with the change in the crystal from (Section 3.10.2), therefore the search models were supplied as two separate ensembles, the PAS-GAF (18 – 312) and the PHY (326 – 503), similarly to how the phase determination of SER26-BV was carried out with two search models derived from the DrBphP and Cph1 structures (Section 3.8.2). The BV was removed from the search model, however, in order to avoid the model bias during the refinement. Molecular replacement was carried out with Phaser either in CCP4 or Phenix, but in either cases phases were determined unambiguously regardless of the space groups chosen from the list of plausible candidates that were originally considered (Table 3.3) wherein molecules were found to form parallel dimers, though the number of asymmetric units vary according to the degree of symmetry of the chosen space group. Structures in space groups of P1 and I422 were refined by cycles of automatic and manual refinement with Refmac5 and Coot. In case of refinement in P1, where eight molecules exist per asymmetric unit, medium non-crystallographic symmetry constraints were applied for all molecules in Refmac5. Refinement in P1 did not reveal any alternative conformation, which suggests that a space group with higher symmetry is correct. The best refinement was achieved in I422 with R-factor/R_{free} values of 0.2773/0.3586, respectively. In comparison to the previous attempts (Section 3.4.3), some improvements can be found. The electron density map covering nearly all the rings of the BV is in accordance with the correct configuration found in the refined mutant structure, albeit some side chains and the thioether linkage with Cys20 is invisible, perhaps due to insufficient refinement, or the radiation damage in case of the latter (Fig. 3.25A). The electron density map at the *Tongue* of the PHY domain which

could not be modelled in the WT preliminary structure due to diffuse density is again, diffuse (Fig. 3.25B).

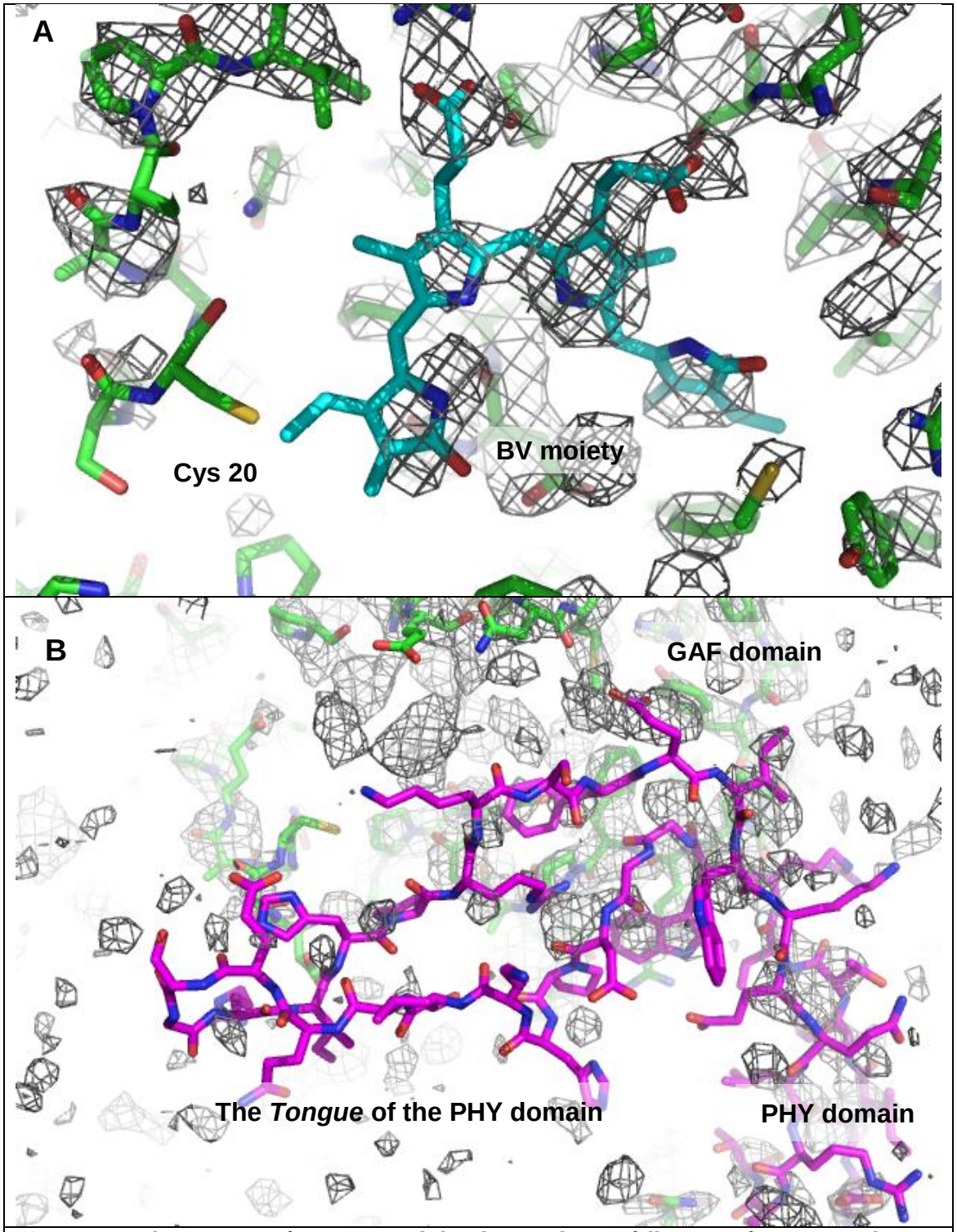


Figure 3.25 The structure of M15-BV and the electron density following refinements with more reliable phase information. The electron density map is contoured at 1.5σ. A) The model and the electron density of the WT M15-BV are shown in green stick and grey mesh, respectively. The electron density covering nearly all the rings of the BV molecule is in agreement with the correct BV orientation, as inferred by the BV model of the refined SER26-BV structure (cyan). B) Only diffuse electron density is found at the *Tongue* of the PHY domain (magenta), similarly to the previous attempt to refine the structure of the WT M15-BV.

3.9.2 Identity of the source of crystallographic difficulties

Even though the quality of the phase information was probably improved by using a more reliable search mode in molecular replacement, the source of crystallographic difficulties in the crystal of the WT M15(s)-BV may still be accounted for by either twinning and/or pseudosymmetry. Relatively low resolution makes the subtle conformational difference between two molecules with apparent crystallographic symmetry very hard to detect, and despite some improvements to the electron density by the more reliable phase information, the WT structure could not be refined satisfactorily within the duration of the present study.

It is noteworthy that structural heterogeneity of phytochromes in one form has been repeatedly documented in the past. Conformational differences were found between monomers in four dimers in the asymmetric unit in the WT structure of PaBphP (3C2W, Yang et al., 2008). A study on Cph1 with NMR spectroscopy, and references therein indicate the heterogeneity of the phytochromes in the Pr-form. Two ground state isoforms were termed Pr-I and Pr-II for Cph1, where these two refer to Cph1 in the Pr-form but differing by their hydrogen-bonding networks and the charge distribution in the tetrapyrrole cavity (Song et al., 2011). Perhaps the unsuccessful refinement of the WT Agp1 structure may be attributed to the natural heterogeneity of phytochromes.

3.10 Structural analysis of Agp1 and comparison to other phytochromes

To the best of the author's knowledge, Agp1 is the first classical bacteriophytochrome to be structurally determined as the entire PCM fragment at high resolution, and this adds to the existing collection of phytochrome structures and enables to correlate the structural features to the underlying determining variables (Table 3.10).

Phytochrome	Spectrophotochromic property	Class	Attached chromophore	Dark stable form
Agp1	Classical	Bacterial	BV	Pr
Cph1	Classical	Cyanobacterial	PCB	Pr
PaBphP	Bathy	Bacterial	BV	Pfr
RpBphP3	Unusual**	Bacterial	BV	Pr
DrBphP	Classical	Bacterial	BV	Pr
SyB-Cph1	Classical †	Cyanobacterial, PAS-less	PCB	Pr
RpBphP1	Bathy	Bacterial	BV	Pfr

Table 3.10 The list of phytochromes whose structures were used in comparative study. **RpBphP3 undergoes photoconversion from the Pr- to Pnr-form (near-red light absorbing state). † The extent to which *SyB*-Cph1 represents canonical phytochromes is still under discussion (Hughes, 2010).

Structural comparisons were carried out mostly between Agp1 and other phytochromes in their WT form in the ground state, because the differences between the WT and the mutant(s) are discussed in the corresponding publications. A high resolution structure of DrBphP bearing a Y307S mutation (PDB code: 2O9C) was exceptionally preferred over the WT structure (PDB code: 1ZTU), however, because of the much higher resolution of the mutant protein structure. With regards to molecular interactions between the BV and amino acid sidechains, Agp1 resemble other phytochromes in the Pr-form. Comparison between Agp1 and the two available Cph1 structures reveals higher similarities between Agp1 and the WT structure (Pr-II isoform, 2VEA) rather than the Y263F mutant structure (Pr-I isoform, 3ZQ5) (Mailliet et al., 2011).

All structural analyses on Agp1 that are described in the following sections are based on SER26-BV structure unless otherwise stated, because the structure of the WT M15-BV could not be refined satisfactorily (Section 3.9).

3.10.1 Domain composition

The structure of Agp1 PCM holoprotein contains PAS, GAF, and PHY domains like the PCM fragments of other phytochromes (Essen et al., 2008, Yang et al., 2008). All three domains in a phytochrome PCM possess a well-established common topology (Essen et

al., 2008, Yang et al., 2008), and the same applies to Agp1 (Fig. 3.26). Each domain, as identified by the SMART server using the WT Agp1 sequence, has an invariable core with five strands forming an anti-parallel β-sheet and an α-helix(ces) between the second and the third β-strands. Additionally, the GAF and the PHY domain possess an α-helix(ces) between the second and the third, and the third and the fourth β-strands. Formation of a trefoil knot, a unique characteristic of phytochromes, is also found in Agp1. The N-terminal region (Asp18-His35) penetrates the loop region between β8 and α8 (Fig. 3.26) of the GAF domain (Pro215 – Ser247), and then interfaces the PAS and the GAF domains. The GAF and the PHY domains are connected by a long helix between α9 of the GAF domain and α10 of PHY domain. The region called the "*Tongue*" or "*Arm*" of the PHY domain between β15 and α15 of the PHY domain (Agp1 Pro 437 – Ser 481) contributes to the inter-domain interface between the GAF and PHY domains (Essen et al., 2008, Yang et al., 2008). The "*Tongue*" of Agp1 consists of two pairs of anti-parallel β-strands, and the distal pair from the PHY domain interacts with a crevice formed by α1, α6, and α8 and thus shielding the BV from solvent access (Fig. 3.27).

Only partial modelling of the α1 helix was possible due to undefined electron densities of the residues preceding Asp 18, even though the expressed protein includes Met 10 to Leu 17 of the M15s construct. The α1 helix is speculated to extend in a similar fashion to Cph1 (2VEA and 3ZQ5) if unhindered (Section 3.10.5). This local disorder is interpreted to be the result of crystal contact between two molecules that do not belonging to one anti-parallel dimer. This crystal contact between two SER26-BV molecules termed arbitrarily as *A* and *B* consists of the tip of the *Tongue* of *A* and β12 in the PHY domain of *B*, and then β18 in the *Tongue* of *A* and α11 in the PHY domain of *B*. There is also an intra-molecular interaction between the tip of the *Tongue* and the N-terminal helix, much like in Cph1 (Essen et al., 2008, Mailliet et al., 2011). If unmodelled residues of SER26 (Met 10 to Leu 17) of the *A* molecule were allowed to continue with the same conformation as in Cph1 (PDB codes: 2VEA and 3ZQ5), then this will cause a steric clash with the *B* molecule (Fig. 3.28). In hindsight, choosing M15s (Agp1 10-504, LE, and hexahistidine) lacking the first nine residues for carrying out SER mutations may have contributed to crystallisation of SER26-BV in this novel form, as additional N-terminal residues could have prohibited crystallisation. Alternatively, this observation could lead to the structure guided optimisation of the

construct by further removing those N-terminal residues without defined electron densities, and that may further improve the diffraction quality and/or the reproducibility of SER26-BV crystal (Section 3.7.3).

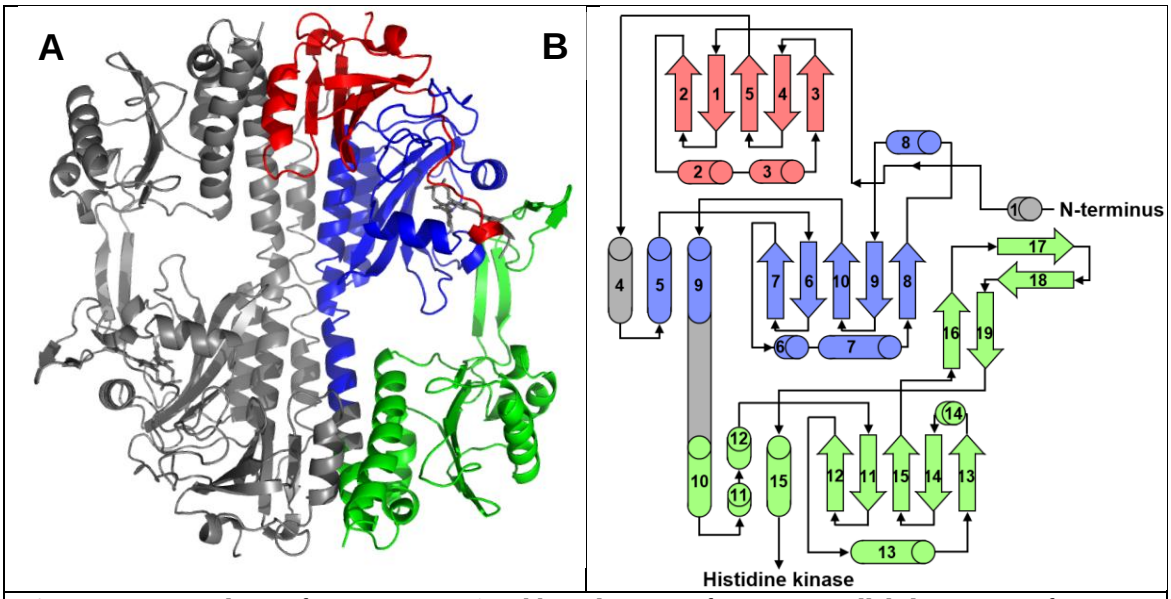


Figure 3.26 Topology of Agp1-M15. A) Ribbon diagram of an anti-parallel dimer pair of SER26-BV molecules. PAS (20-117), GAF (142-320), and PHY (326-504) domains identified by SMART server with WT Agp1 M15 sequence are coloured red, blue, and green, respectively. B) Topology diagram of the Agp1 structure. α-helices and β-strands are represented by cylinders and arrows, respectively. α-helices and β-strands are separately numbered in order from the N-terminus to the C-terminus. Figure adapted and modified from Yang et al., 2008.

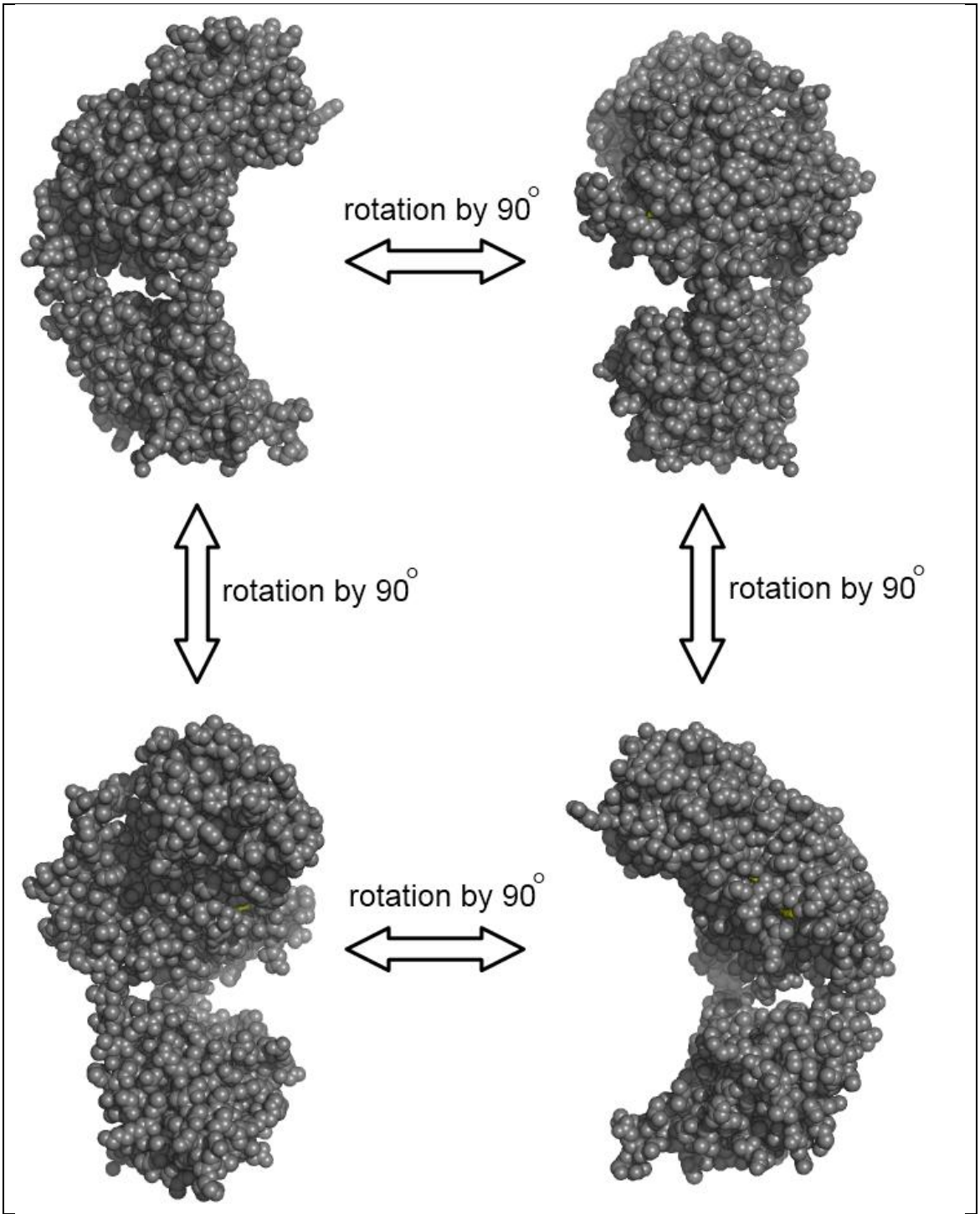


Figure 3.27 The BV molecule is shielded from the solvent in the Agp1 holoprotein structure. Four structures are views from four angles that are 90 ° apart along the y-axis. Structures are shown in space-filling model with the protein in grey and the BV in yellow. The BV is hardly visible from all directions, and this is an indication of effective solvent shield.

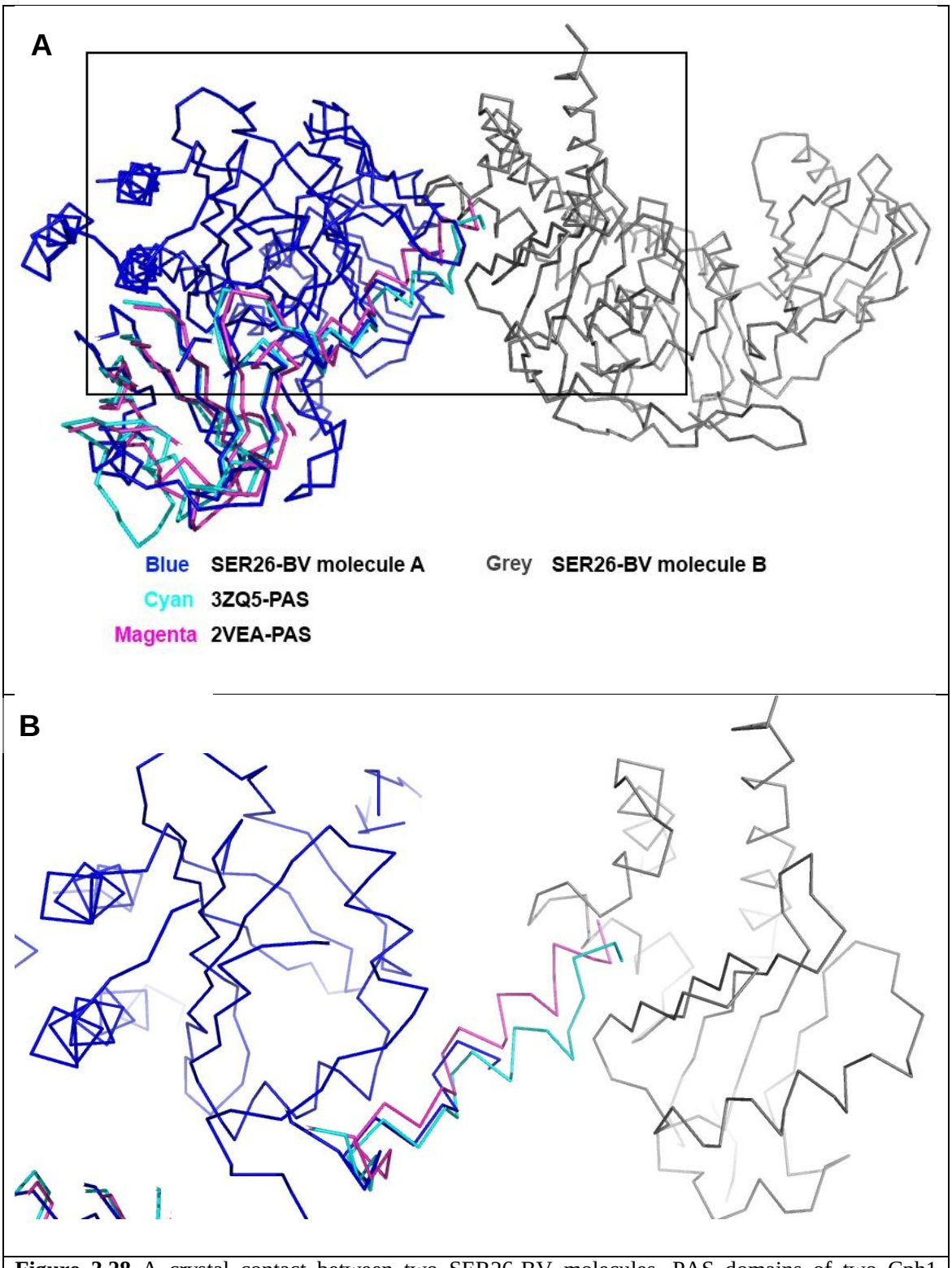


Figure 3.28 A crystal contact between two SER26-BV molecules. PAS domains of two Cph1 structures were truncated to Val5, the equivalent of the first residue of SER26 construct (Met10), and superposed onto the SER26-BV structure. A) An overview of the crystal contact site. B) Close-up view of the enclosed region in the overview. The PHY domain of the SER26 A molecule is omitted for clarity. Structural superposition between SER26-BV A molecule and PAS domains of Cph1 (2VEA and 3ZQ5, Val5 – Ala125 was achieved with global superposition with SER26-BV as the reference.

3.10.2 Inter-domain orientation and the mode of dimerisation

SER26-BV crystallised as anti-parallel dimer rather than a parallel dimer as in WT M15-BV. Other phytochrome PCMs have crystallised as anti-parallel dimer, but furthermore SER26-BV and the crystal structure of the WT Cph1 (PDB code: 2VEA) coincidentally share the same space group with one molecule per asymmetric unit, therefore the unit cell parameters are similar as a result (Table 3.11). Smaller unit cell volume by SER26-BV (SER26-BV, 113911.72 $Å^3$; 2VEA, 1491643.90 $Å^3$) is a consequence of its more compact conformation arising from a tight bend at the linker helix between the GAF and the PHY domains. The angle between the first and the last ten residues of the linker helix between the GAF and the PHY domains in M15-BV (WT), SER26-BV, and Cph1 (2VEA) are 13.89 °, 49.12 °, and 29.66 °, respectively (Fig. 3.29). A tight bend by SER26-BV at the linker helix between the α9 of the GAF and the α10 of PHY domain has the highest angle, and this further confirms the flexibility in this region as discussed in previous publications (Essen et al., 2008, Mailliet et al., 2011, Yang et al., 2008, Yang et al., 2009). There is a weak correlation between the form of phytochromes and the angle at the linker helix. Phytochromes in the Pfr-form tend to adopt more straight conformation and vice versa. The 3ZQ5 structure of Cph1 in the Pr-form with a straight linker helix does not fit to this rule, however, but crystal structures are subject to packing induced conformational change. More structural data of phytochromes are required to enable identification of conformational outliers. Alternatively, electron microscopy and/or SAXS may provide structural information that better represents the global conformation *in vivo*.

SER26-BV crystallised as anti-parallel dimer with a larger buried surface area at the dimer interface (1960.7 $Å^2$) compared to the parallel dimers in the WT M15-BV crystal (1374.5 $Å^2$). The dimerisation of the WT M15-BV is mediated by helical bundles involving N-terminal and C-terminal regions of the linker helix between the GAF and PHY domains. The helical bundles at the N-terminus of the linker helix involve α4 and α9 of each monomer whereas the helical bundles at the C-terminus of the linker helix involve α10 and α15 of each monomer. The dimer interface of SER26-BV mainly consists of α10 of the PHY domain of a molecule interacting with the α9 and α4 of the symmetry related dimer mate. The anti-parallel dimer of SER26-BV resembles that of the two crystal structures of Cph1 (2VEA and 3ZQ5). On a closer inspection, there is a subtle difference in anti-parallel structures of Cph1 and SER26-BV. In Cph1 the dimers

appear to be more elongated because the monomers are somewhat more offset from the twofold symmetry axis (Fig. 3.30).

All the crystal structures of phytochromes solved to date form dimers, but the mode of dimerisation is not uniform (Table 3.11). As described in Section 1.2.4, all SKs must form a parallel dimer to carry out its kinase activity, thus the anti-parallel dimerisation is believed to be an artefact of crystallisation and does not reflect the quaternary structure *in vivo*.

The factors that determine the mode of dimerisation in phytochrome PCM structures were proposed to be either of the following – the crystallisation condition or the lineage of the phytochrome. Prior to crystallisation of SER26-BV, anti-parallel dimerisation was only observed in crystals of Cph1 (cyanobacterial phytochrome) obtained under crystallisation conditions of high ionic strength. Both assumptions were invalidated by the observation that crystallisation of SER26-BV (bacterial phytochrome) as an anti-parallel dimer occurs in at low ionic strength (Fig. 3.30).

Recently the structure of RpBphP2 CBD was determined (Bellini and Papiz, 2012a). Since the WT RpBphP2 proved recalcitrant to being crystallised, an extensive effort was made for its crystallogenesis by replacing the residues corresponding to those participating in crystal contacts in the crystal structure of RpBphP3 (Yang et al., 2007). Therein, the possibility that an Arg residue (RpBphP3 Arg 150) plays a key role in mediating parallel dimer in phytochromes was proposed. This is in accordance with the crystallisation propensities of Agp1 M20 and M15 holoproteins which is discussed in Section 3.11.1 (See also Section 6.3).

The factor which determines that mode of dimerisation was attributed to the presence of epitopes that can mediate crystal contacts. In Section 5.5, a study is presented where correlations between the presences of such epitopes and possible crystal forms are shown. Our crystallographic studies on Agp1 M15 proteins have shown that it is possible for a phytochrome PCM to crystallise as both parallel and anti-parallel dimer, therefore it is plausible that the subtle structural differences at the protein surface, rather than other physico-chemical factors, determine whether one or another mode of dimer formation occurs in a crystal.

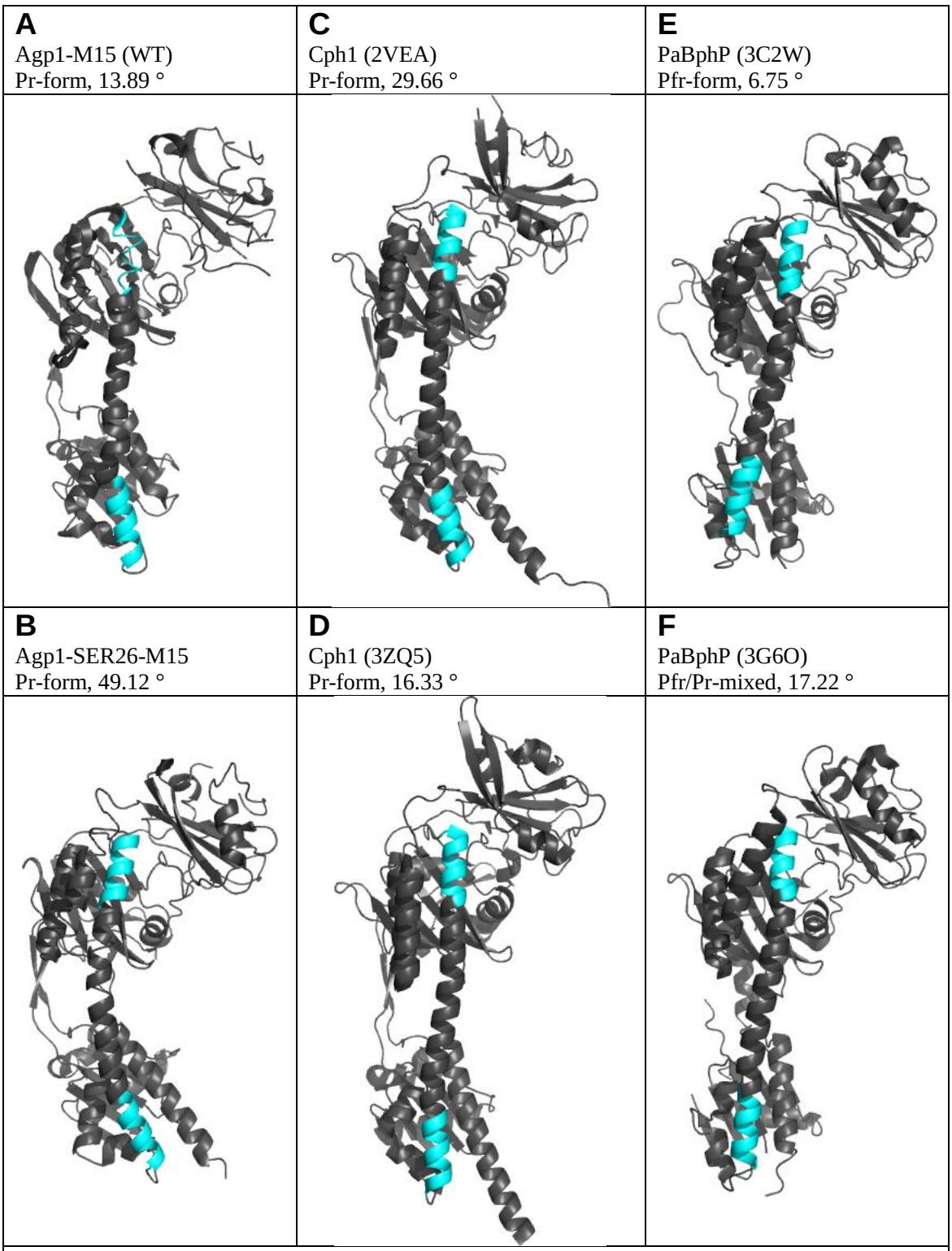


Figure 3.29 Global conformations of phytochrome PCM structures. Structures were superposed globally and are presented in the same orientation. The first and the last ten amino acid residues used to calculate the bend of the inter-domain helix are coloured cyan. The proteins (PDB codes) and the residue ranges are as follows. A) Agp1 WT preliminary structure, 289-298 & 325-334. B) Agp1 SER26-BV, 289-298 & 325-334. C) WT Cph1 (2VEA), 298-307 & 336-345. D) Cph1 Y263F (3ZQ5), 298-307 & 336-345. E) WT PaBphP (3C2W), 286-295 & 318-327. F) PaBphP Q188L (3G6O), 285-294 & 321-330. The 3IBR structure of PaBphP Q188L was omitted as the global conformation is essentially identical to that of the 3G6O structure.

Year	Phytochrome	Fragment	PDB code	Mode of dimerisation
2005	DrBphP	CBD	1ZTU	Parallel
2007	RpBphP3	CBD	2OOL	Parallel
2007	DrBphP Y307S	CBD	2O9C	Parallel
2008	Cph1	PCM	2VEA	Anti-parallel
2008	PaBphP	PCM	3C2W	Parallel
2009	PaBphP Q188L	PCM	3G6O, 3IBR	Parallel
2011	Cph1 Y263H	PCM	3ZQ5	Anti-parallel
2012	DrBphP D207H	CBD	3S7N, 3S7O, 3S7P	Parallel
2012	DrBphP D207H/Y263F	CBD	3S7Q	Parallel
	Agp1	PCM	N/A	Parallel
	Agp1 E86A/E87A/E336A/K337A	PCM	N/A	Anti-parallel

Table 3.11 List of crystal structures of phytochromes and the mode of dimerisation in chronological order. All CBD fragment crystals described so far form parallel dimers, but both parallel and anti-parallel dimers are found for PCM fragments.

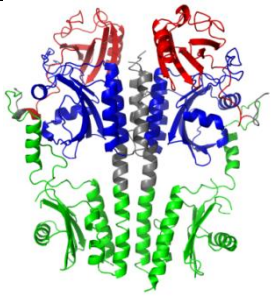
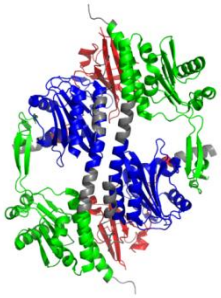
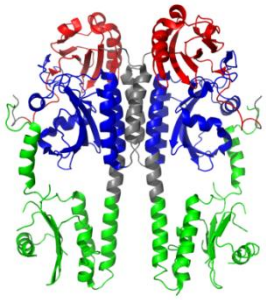
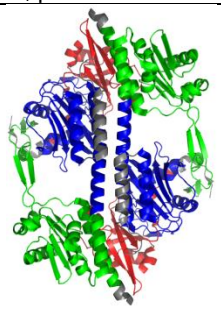
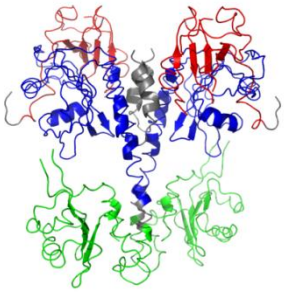
Parallel	*Anti-parallel*
A  **PaBphP (3C2W)** 0.45 M $(NH_4)_2SO_4$, 0.1 M Tris-HCl, pH 7.7	**D**  **Cph1 (2VEA)** 2 M Na Acetate, 0.1 M Mg Acetate, pH6.7
B  **PaBphP Q188L (3G6O, 3IBR)** 0.5% PEG4000, 0.01 M Na Acetate, pH 4.6	**E**  **Cph1 Y263H (3ZQ5)** 2.5 M Na Acetate, 0.1 M MES pH 6.5
C  **Agp1 M15-BV** 200 mM $CaCl_2$, 16% PEG3350, 0.1 M Tris-HCl, 2.5% Glycerol, pH 8.5	**F**  **Agp1 SER26 (E86A/E87A/E336A/K337A)** 0.03 M diethyleneglycol, 0.03 M triethyleneglycol, 0.03 M tetraethyleneglycol, 0.03 M pentaethyleneglycol, 20% Glycerol, 10% PEG8000 0.05 M MES, 0.05 M Imidazole, pH 6.5

Figure 3.30 Mode of dimerisation and crystallisation conditions of phytochrome PCM crystal structures. Parallel dimer exhibited by A) PaBphP (3C2W), B) PaBphP Q188L (3G6O, 3IBR), C) Agp1. Anti-parallel dimer exhibited by D) Cph1 (2VEA), E) Cph1 Y263H (3ZQ5), F) Agp1 E86A/E87A/E336A/K337A. Ribbon diagrams are coloured according to the domains as defined by SMART server. Agp1 (PAS, 20-117; GAF, 142-320; PHY, 326-504), PaBphP (PAS, 12-113; GAF, 139-313; PHY, 314-498), Cph1 (PAS, 15-125; GAF, 152-330; PHY, 331-513). Red, PAS; Blue, GAF; Green, PHY.

3.10.3 Electrostatic surface potentials of phytochrome PCMs

The electrostatic surface potential of Agp1 was calculated with the PDB2PQR server followed by APBS (Section 2.2.8.4) along with the PCMs of other phytochromes in their ground state (Fig. 3.31). Distributions of positive and negative charges as well as non-polar regions at the dimer interface indicate that the modes of dimerisation for PCM of Agp1, Cph1, and PaBphP are mixtures of ionic and hydrophobic interaction, and cannot easily be attributed to a single type of interaction.

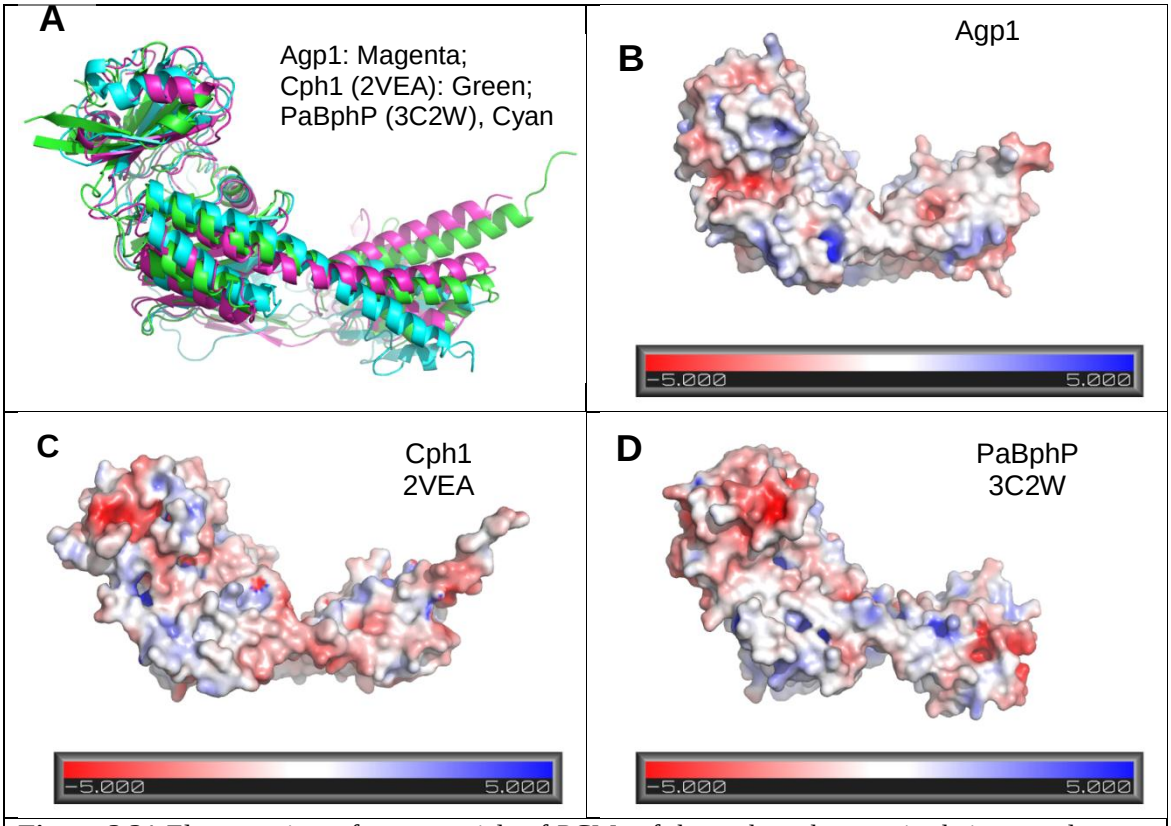


Figure 3.31 Electrostatic surface potentials of PCMs of three phytochromes in their ground states. A) Structures of three phytochromes were superposed globally and presented in ribbon diagram and orientated for viewing the dimer interface. Agp1, Magenta; Cph1 (2VEA), Green; PaBphP (3C2W), Cyan. APBS electrostatic surface potential diagrams of B) Agp1, C) Cph1, and D) PaBphP, coloured by electrostatic potential (-5 kT/e, red to +5 kT/e, blue).

3.10.4 Determination of the mode of biliverdin binding

As described in Section 1.3.2, a correlation exists between the types of the bound bilin and the location of a conserved cysteine where the covalent bond occurs, i.e. plant and cyanobacterial phytochromes bind to PΦB and PCB, respectively, via a cysteine in the GAF domain, whereas bacterial and fungal phytochromes bind to BV with a chromophore-bonding cysteine in the N-terminal of the PAS domain. PΦB/PCB and BV feature ethylidene and vinyl side chains at C3 of the A-ring, respectively, in their free form (Fig. 1.7). The geometry of the C2 atom of the bilin chromophore could change following the assembly of the holoprotein, therefore the orientation of the methyl side chain at the C2 atom could give information regarding the mode of binding between the bilin and the phytochrome.

In the first phytochrome structure (1ZTU), the A-ring of the BV was modelled with an endocyclic double bond, therefore the methyl side chain at the C2 atom was assumed to maintain the co-planar geometry to the A-ring (Fig. 3.32A) (Wagner et al., 2005). Hereafter the term "co-planar form" will refer to the geometry of the A-ring methyl side chain of the bilin following the originally proposed mode of BV attachment.

The A-ring of PΦB (and PCB) adopts a non-planar geometry with an exocyclic double bond and the C2 atom assumes S chirality upon formation of the covalent bond with plant (or cyanobacterial) phytochrome (Fig. 3.32B), except for the PCB of *SyB*-Cph1 which has the R chirality at the C2 atom (Cornilescu et al., 2008).

A new BV attachment mechanism was proposed for bacteriophytochromes when a detailed analysis was carried out with the high resolution structure of DrBphP (Wagner et al., 2007). The $C2^1$ methyl carbon of the BV rotates towards Met 259 in the high resolution structure, hence the C2 carbon was deduced to assume a tetrahedral configuration. Consistently the BV A-ring was modelled with exocyclic double bond and non-planar geometry and the BV was proposed to convert to 2(R),3(E)-PΦB following the assembly process, which has the opposite chirality at the C2 carbon of the bilin of plant/cyanobacterial phytochromes (Fig. 3.32C).

It must be noted, however, that the DrBphP structure was solved in a truncated form of CBD, therefore leaving a possibility for an inadvertent artefact to be present. Since the diffraction dataset for Agp1 SER26-BV in the present study extends to 1.82 Å

resolution, the highest resolution reported for crystals of a phytochrome PCM to date. Geometric analysis of the A-ring was attempted.

Geometric analysis was performed by carrying out independent refinements where the geometry of the BV is constrained according to the three forms. The libraries that define the geometric constraints during refinements were generated by extracting the coordinates of the BV molecule from the high resolution DrBphP structure, and the coordinates were modified with Discovery Studio 3.1 (Accelrys). Three models were generated with the methyl side chain at the C2 atom of the A-ring in two chiralities (S and R) or co-planar with the A-ring, therefore the geometries were the same as those in holoproteins depicted in Fig. 3.32 B, C, and A, respectively. Three library files with different configuration at the C2 carbon were obtained as described (Section 2.2.7.2.4, Section 3.8.2). The library files were used during refinement with Refmac5 (Murshudov et al., 2011), where the weight matrix was either defined automatically by Refmac5 or as an arbitrarily small value of 0.1 (Section 2.2.7.2.3). In all three geometric modes, automatically defined weighting terms were 0.5. The output coordinates, the electron density map (2*Fo* - *Fc*), and the difference maps (*Fo* - *Fc*) were inspected.

Firstly, a possibility for the tetrahedral C2 atom to be in S-form like plant/cyanobacterial phytochrome (S-form) was immediately ruled out. When the refinement was carried out with the C2 atom in S-form with automatically defined weighting, the methyl side chain was placed in a rather planar orientation (Fig. 3.33B), which was suspected to be the result of counteracting effects of geometrical constraints defined by the library and the $C2^1$ atom being moved to fit into the electron density map. The hypothesis was confirmed when the refinement was repeated with weighting value of 0.1, where the orientation of the methyl side chain was predominantly determined by the library. A positive density for the methyl side chain appeared to indicate that either the co-planar configuration or R-forms to be correct (Fig. 3.33C).

It is more difficult to judge from 1.82 Å resolution data whether the stereochemistry of the C2 atom is in the non-chiral or R-form, however. When the refinement was carried out with automatically determined weighting, the $C2^1$ atom resulted in a good fit into the electron density map regardless of the stereochemistry of the C2 atom. Though data collected to 1.82 Å is not sufficient by its own to draw a decisive conclusion, the electron density map of Fig. 3.33 is slightly skewed towards R-from of C2 atom. Additionally, the double bond between the C3 and $C3^1$ atom of the A-ring is co-planer

to the A-ring (Fig. 3.32C) in the refined Agp1 structure, and this suggests the presence of exocyclic double bond. Therefore the geometry of the bilin molecule was modelled identically to the BV in DrBphP structure which features an exocyclic double bond with C2 atom in the R-form.

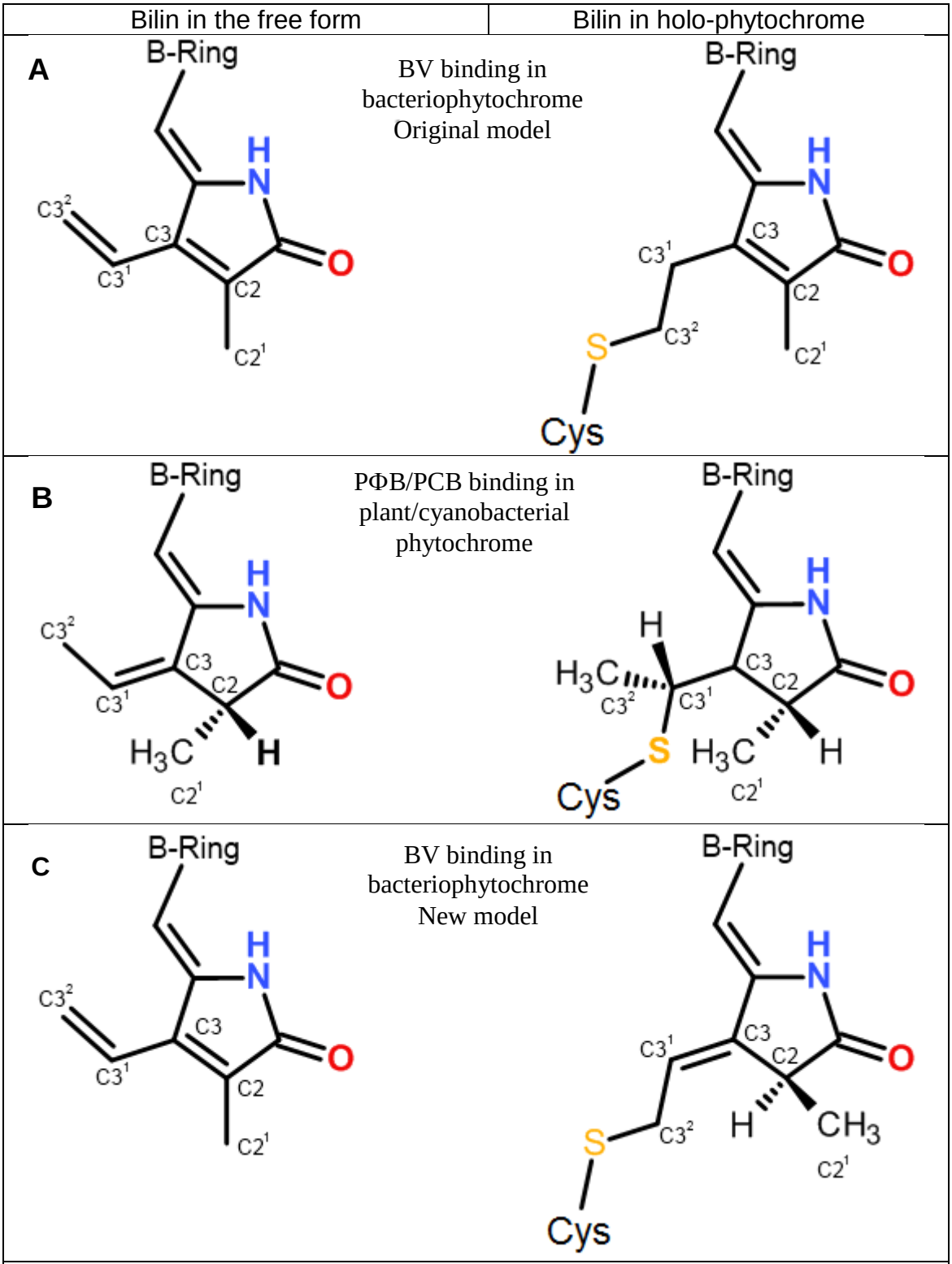


Figure 3.32 Formation of the covalent bond between the cysteine and the bilin chromophores. Structure of the A-ring of the bilin in the free form (Left) and following the lygation (Right). A) Originally proposed mode for bacteriophytochrome in which a thioether bond forms between $C3^2$ atom of the BV and the S-gamma atom of the conserved Cys residue in the N-terminal of the PAS domain, and the A-ring maintains the endocyclic double bond. B) The mode for plant and cyanobacterial phytochromes, in which tetrahedral C2 atom of the PCB adopts the S-form. C) A binding mode for bacteriophytochrome based on the high resolution structure of DrBphP (Wagner et al., 2007). BV converts to 2(R), 3(E)-PΦB, with its tetrahedral C2 atom in R-form and a non-planar A-ring with exocyclic double bond. Figures adapted from Wagner et al., 2005 (A) and Wagner et al., 2007 (B, C).

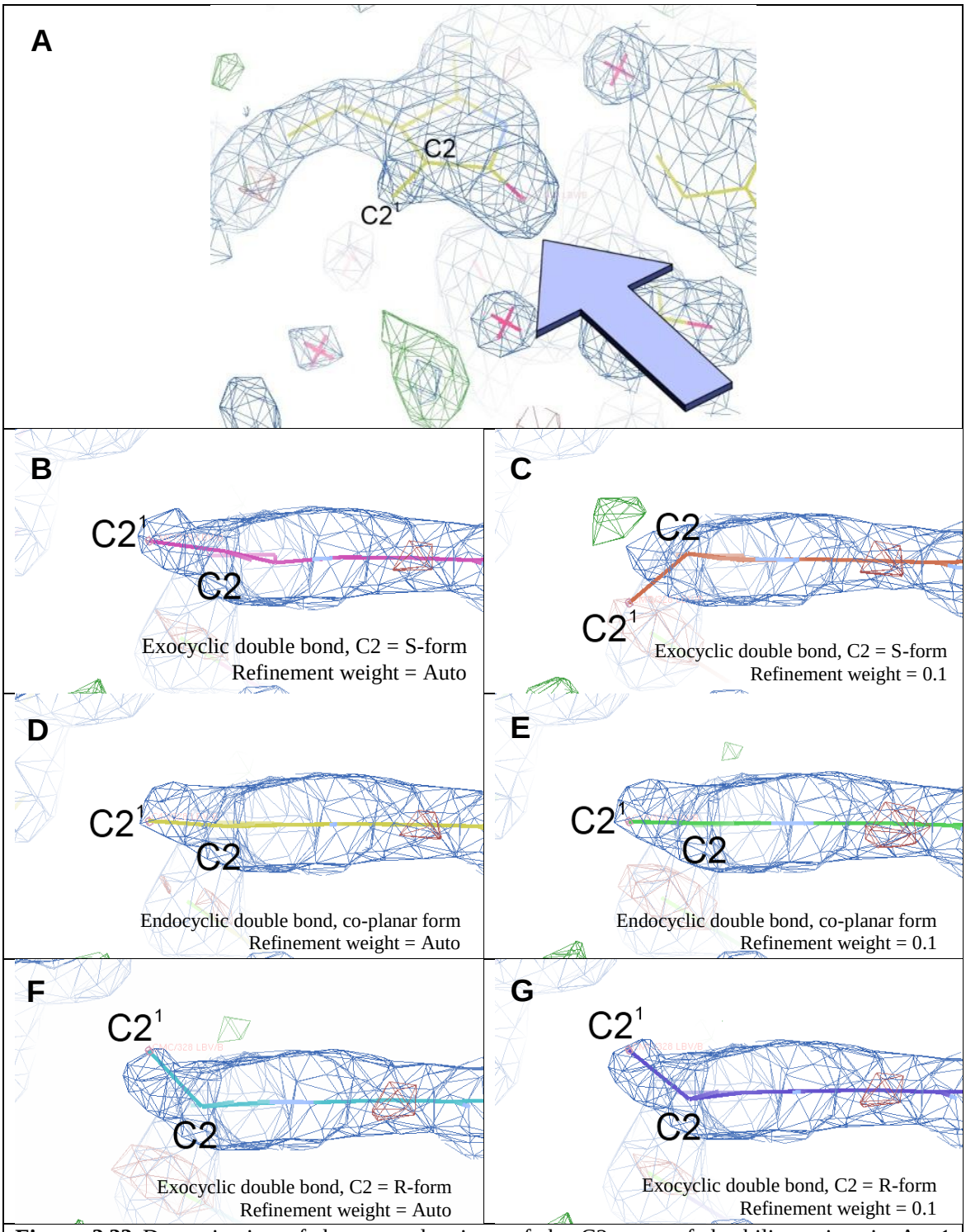


Figure 3.33 Determination of the stereochemistry of the C2 atom of the bilin moiety in Agp1 structure. A) The arrow indicates the point of view for the subsequent diagrams. B) C) Results following a refinement with C2 atom of the bilin in S-form with weighting of automatic and 0.1, respectively. D) E) Results following a refinement with C2 atom of the bilin in the co-planar form with refinement weight values of automatic and 0.1, respectively. F) G) Results following a refinement with C2 atom of the bilin in R-form with refinement weight values of automatic and 0.1, respectively. Automatically determined weighting terms were 0.5 in all three geometric modes.

3.10.5 The N-terminal region

Two α-helices in the vicinity of the bilin pocket were found to be collinear in the WT Cph1. Cph1 α1 and α7 are located at the N-terminal region and the GAF domain, respectively (Essen et al., 2008). Within the same study, bacteriophytochromes structures available at that time (DrBphP, Wagner et al., 2005, Wagner et al., 2007; RpBphP3, Yang et al., 2007) were described as lacking the α-helix in the N-terminal region which was interpreted as due to either the lack of PHY domain or the presence of the chromophore-binding Cys residue in the N-terminal region. Structures of N-termini of the WT and Q188L mutant of PaBphP were interpreted as unstructured coil, which was described as a contrast to a three-turn helix of the Cph1 structure by Yang et al. (2009).

Analysing phytochrome structures with DSSP, however, presents more complete picture. An α-helix was found in the N-terminus of Agp1 from Ser 19 to Ala 22 (Fig. 3.34). Furthermore, 3/10 helix was found at the N-terminus of DrBphP (2O9C), RpBphP3 (2OOL), and PaBphP (3C2W), but not in structures of Q188L mutant of PaBphP (3IBR and 3G6O).

An α-helix in the N-terminal region of Agp1 refutes one of the hypotheses presented by Essen et al. (2008), because Agp1 is a bacteriophytochromes with the chromophore-binding Cys residue in the N-terminus. Conversely, the other hypothesis of attributing the lack of the N-terminal α-helix to the lack of the PHY domain is not refuted. Phytochrome CBD can form the Pr-form as defined spectrophotometrically, but it is still plausible to hypothesise that the PHY domain to influence the formation of the N-terminal region on the basis of close proximity of the *Tongue* of the PHY domain and the N-terminal region as seen in phytochrome PCM structures.

The helix of Agp1 is shorter than that of Cph1 structures, and the reason for the lack of electron density in the N-terminal region was speculated in Section 3.10.1 (Fig. 3.28). It also remains to be seen for other bacteriophytochromes if the N-terminal α-helix is present and if so, it will be as long as that of Cph1 if unperturbed by a crystal contact.

Whether the 3/10 helix at the N-terminus of PaBphP is its *bona fide* feature or not is difficult to discern, because the helix was not found in Q188L mutant structures in mixed state. It is possible that the conformation of this region was assuming 3/10 helix by chance. Currently PaBphP is the only phytochrome with the structure of the Pfr-form,

therefore the sample size is too small to allow confident description of the structure of the Pfr-form in general.

Since the structural determination of PCM of Cph1 (Essen et al., 2008) and PaBphP (Yang et al., 2008), there has been a hypothesis that the differences in secondary structure between them to reflect the differences between the Pr- and Pfr-forms. CD-spectroscopy was employed to investigate this hypothesis and the result is described in Section 6.6.2.

The structure of Agp1 and detailed analyses on other structures enable more accurate deduction of the underlying variables which are manifested as the structure, but more structures especially of the Pfr-form will enhance our understanding.

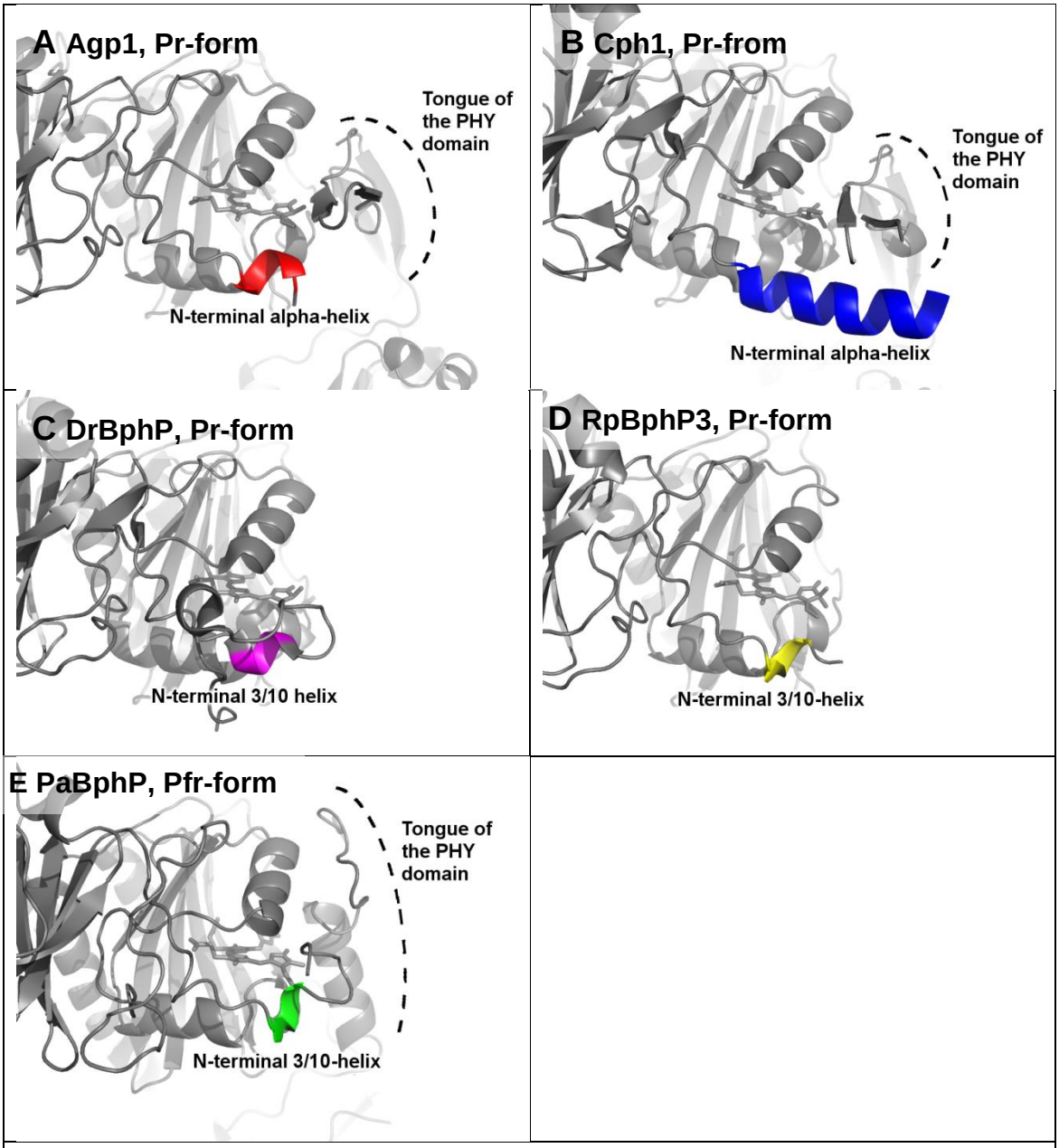


Figure 3.34 Conformations at the N-terminal region preceding the PAS domain. Helices in the N-terminal region as defined by DSSP are coloured. A) Agp1 SER26-BV, B) Cph1 (2VEA), C) DrBphP (2O9C), D) RpBphP3 (2OOL), E) PaBphP (3C2W). 3/10 helix in PaBphP was not rendered as a helix by default in PyMOL, therefore the secondary structure assignment was manually changed. Structures of mutants were checked for their secondary structures at the N-terminal region, and were found to be essentially the same as the wild type for Cph1 (WT: 2VEA and mutant: 3ZQ5) and DrBphP (WT: 1ZTU and mutants: 2O9C, 3S7N, 3S7O, 3S7P, and 3S7Q). Structures of PaBphP (WT: 3C2W and mutants: 3IBR, and 3G6O) were similar, but 3/10-helix was only found in the WT structure by DSSP.

3.10.6 The PHY domain

The PHY domain of phytochromes are characterised by the presence of the *Tongue*, an extension that interacts with the GAF domain. The *Tongue* of Agp1 consists of two anti-parallel β-sheets similar to that of Cph1, except that the tip is shorter for Agp1. This contrasts with that of PaBphP, featuring a coil and an α-helix (Fig. 3.35). This conformational difference between the Pr and the Pfr-form is believed not to be an artefact of crystal packing, because the secondary structures at the *Tongue* are conserved between the WT and the mutant structures despite the differences in the crystal form (Table 3.12). The preliminary structure of the WT Agp1 also has the *Tongue* with a β-sheet structure as far as the electron density of limited quality can be interpreted, despite crystallisation in a different crystal form than SER26-BV.

Similarities between Agp1 and Cph1 are also found within the core of the PHY domain. The loop prior to α13 in Agp1 and the equivalent structure in Cph1 share a similar conformation, which is distinct from that of PaBaphP. In classical phytochromes the loop conformation is partially due to the presence of a conserved Pro (Pro 376 in Agp1, Pro 383 in Cph1), which is absent in PaBphP (Fig. 3.36). This Pro interacts hydrophobically with Tyr 412, Val 415, and Phe 435 (corresponding residues in Cph1 are Phe 419, Val 422, and Phe 440).

Two short helices immediately following the GAF-PHY linker helix are found in Agp1 and Cph1, and in Agp1 the angle between α11 (Trp 338 – Gly 344) and α12 (Glu 346 – Val 356) is 111.8°. Only one helix corresponding to the former one is found in PaBphP, the rest of the region being occupied by a loop.

It is difficult to comment on the effect of these conformational differences between Pr and Pfr-forms in terms of signal transduction in the absence of a full-length structure of a phytochrome.

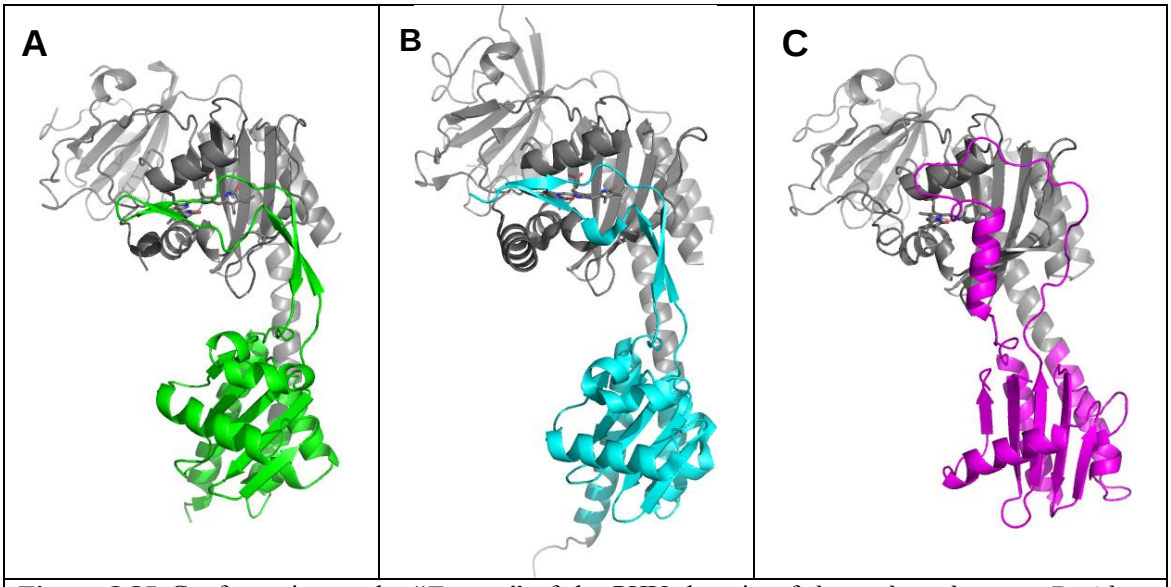


Figure 3.35 Conformation at the "*Tongue*" of the PHY domain of three phytochromes. Residues corresponding to 326 to 504 of Agp1 are coloured. A) Agp1 SER26-BV, B) Cph1 (2VEA), C) PaBphP (3C2W). Phytochromes in the Pr-form possess β-strands in the region, whereas the corresponding region of PaBphP in the Pfr-form comprises coil and an α-helix.

Phytochrome	**Variation**	**PDB code**	**Space group**
PaBphP	WT	3C2W	$C222_1$
PaBphP	Q188L mutant	3BIR and 3G6O	$P6_5$
Cph1	WT	2VEA	$P4_32_12$
Cph1	Y263F mutant	3ZQ5	C121

Table 3.12 List of phytochrome PCM crystal structures and their space groups. Space group of the WT and the mutant crystals are different, but secondary structures are well conserved.

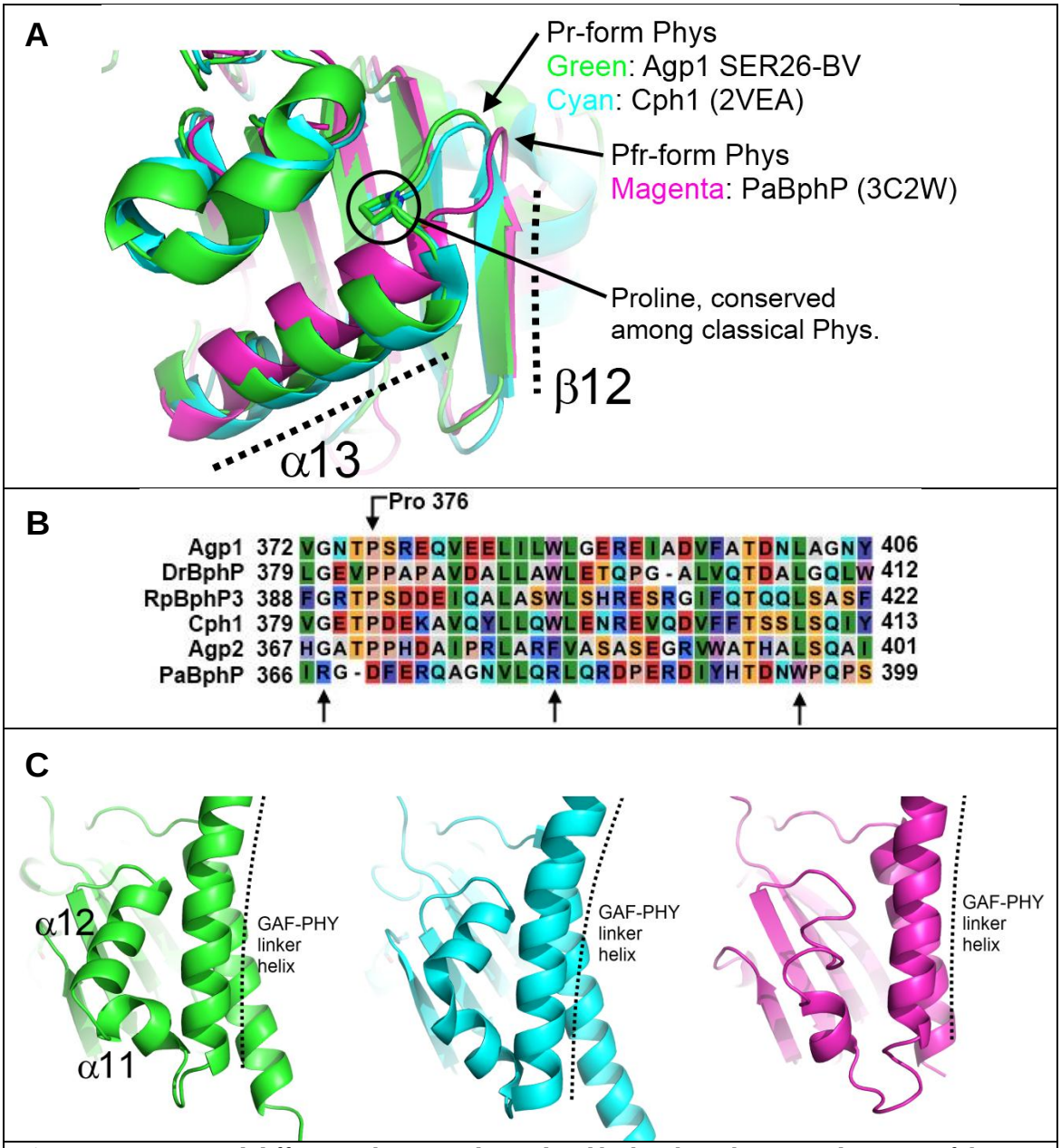


Figure 3.36 Structural differences between classical and bathy phytochromes. Alignment of the core of the PHY domain was achieved with LSQ superpose in Coot with the following range of residues. Agp1 (Asp 359 – Arg 436), Cph1 (Gly 367 – Arg 441), and PaBphP (Asp 353 – Arg 428). A) A conserved Pro residues in classical phytochromes contributes to conformation of the loop between β12 and α13, which is different between classical (Agp1 and Cph1) and bathy (PaBphP) phytochromes. B) Sequence alignment indicates some unique amino acid contents by PaBphP unshared with Agp2, another bathy-phytochrome (indicated by arrows). Instances of non-conserved residues unique to PaBphP accord with a hypothesis of bathy-phytochromes to have multiple origins (Rottwinkel et al., 2010). C) Conformational differences at the helix(ces) that immediately follow(s) the GAF-PHY linker helix. Green: Agp1, Cyan: Cph1, Magenta: PaBaphP.

3.10.7 The Bilin chromophore

The BV molecule of Agp1 has the same stereochemistry as that of DrBphP, which is 5Zs10Zs15Za. Overall, Agp1 exhibits chromophore-protein interactions conserved in other phytochromes in the Pr-form, such as DrBphP, RpBphP3 and Cph1. With regards to the two Cph1 structures, Agp1 is more similar to 2VEA (proposed isoform II) than 3ZQ5 (Y263F mutant, proposed isoform I) in terms of side chain orientation in the bilin pocket. Structural features where the RpBphP3 structure deviates from the norm were regarded to reflect its unique photochromic property to give rise to the Pnr-form upon irradiation at λ_{Qmax}.

The D-ring of the BV in Agp1 is tilted with respect to the co-planar B and C-rings, much like in other phytochrome structures in the Pr-form., but more similar to bacteriophytochromes (RpBphP3: 2OOL and DrBphP: 2O9C) than cyanobacterial phytochrome (Cph1: 2VEA). In Agp1, the angle between the D-ring and the C-ring is 50.5 °, which is more similar to bacteriophytochromes (DrBphP, 1ZTU – 44.4 °, 2O9C - 50.8 °; RpBphP3: 2OOL - 39.8 °) than a cyanobacterial phytochrome (Cph1: 2VEA – 26.3 °, 3ZQ5 – 37.2 °). The conformation of the BV in Agp1 is a result of conserved hydrophobic pocket consisting of Y166, V176, F193, P194, Y253, and F465 in the vicinity of the D-ring.

Agp1 is unique among phytochromes in the Pr-form in that the relative position of the conserved His 280 is on the β-facial side of the D-ring, rather like PaBphP in the Pfr-form. The importance of this His in the formation of the Pfr-form has been demonstrated in both classical and bathy phytochromes (Zienicke et al., 2010, Wagner et al., 2008, Yang et al., 2009). Also Agp1 stands out from all other structurally determined phytochromes in terms of the tilt of the imidazole ring of His 280 relative to the D-ring, because the direction is opposite (Fig. 3.38). Nevertheless a hydrogen bond between the NE2 atom of the conserved histidine and the oxygen atom of the D-ring of the bilin is conserved between Agp1, DrBphP, and Cph1 (Fig. 3.39 and Fig. 3.41).

This unusual side chain conformation is brought about by polar as well as hydrophobic environments surrounding the His 280 that are unique to Agp1, and there are several deviations in the corresponding segments of amino acid sequences in Agp1 from the other classical phytochromes (Fig. 3.37 and 3.39).

Firstly, in Agp1 a non-conserved Leu 164 occupies the space of the typical Met residue (DrBphP: Met 174, Cph1: Met 174). This qualitatively similar but shorter side chain makes hydrophobic interaction to the D-ring just like Met residues in other phytochromes (Fig. 3.39), but may allow the BV or surrounding residues to adopt slightly different conformation by the virtue of its smaller dimension. The presence of a positively charged lysine in this position in RpBphP3 and *SyB*-Cph1 may reflect their unique photochromic activity. Secondly, a position occupied by hydrophobic residues in other phytochromes is occupied by polar Thr 259, and this creates a water mediated hydrogen bond between Thr 259 and His 280, which is absent in other phytochromes. Thirdly, the hydrophobicity at the position of Met 254 is conserved, but the side chain conformation is different as it points away from His 280. Additionally, the conformation of His 280 may be influenced by Arg 162. The conformation of the latter is only conserved among Agp1, RpBphP3, and one of two alternative conformations in the Cph1 2VEA structure, but different from DrBphP (2OOL), Cph1 (the other of two alternative conformations in 2VEA and 3ZQ5). The conformation of Arg 162 in Agp1 may be determined by two amino acids in its vicinity, Leu 185 and Pro 186, neither of which are spatially well conserved. The conformation of RpBphP3 His 299, the equivalent of His 280 in Agp1, is similar to other phytochromes despite the presence of Arg 181 being in a conformation similar to that of Agp1. This may be attributed to the presence of a non-conserved Lys 183 to prevent the tilt of imidazole ring of His 299.

It was proposed in Essen et al., 2008 that greater tilts observed at the D-ring of BV in CBD fragments of bacteriophytochromes may be an artefact caused by the lack of the PHY domain. However, the structure of the entire PCM of Agp1 suggests as one possible scenario that the high D-ring tilt is the norm for bacteriophytochromes. Alternatively, it is possible that the high D-ring tilt in Agp1 even in the presence of the PHY domain is unique to Agp1, caused by an unusual conformation of His 280. More structural evidence is required to fully prove the structural isomorphism between CBD and PCM. For example, knowledge of the entire PCM structure of DrBphP would help to clarify this issue.

As of now it is difficult to rationalise how both conserved and non-conserved interactions involving His 280 may affect the spectrophotochromic properties of Agp1. Hydrogen bond between the His residue and the oxygen atom of the D-ring is conserved, but under an assumption of structural isomorphism of CBD and PCM, there is no clear correlation between the λ_{Qmax} of the Pr-form and the tilt angle of the D-ring in

bacteriophytochromes (Agp1 702 nm, 50.5 °; RpBphP3 705 nm, 2OOL - 39.8 °; DrBphP 700 nm, 1ZTU - 44.4 °, 2O9C - 50.8 °). SAR values (A_{Qmax}/A280) have not been reported for RpBphP3 and DrBphP, therefore no comparison could be made to Agp1 in this regard.

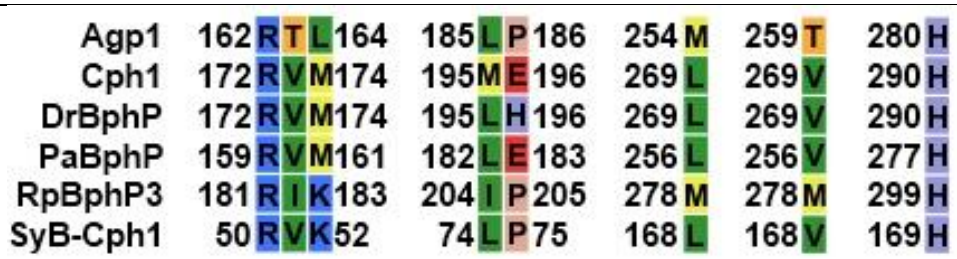


Figure 3.37 Sequence alignment of amino acid residues that interact with His 280 in Agp1. Some deviations are found in amino acid sequence in Agp1.

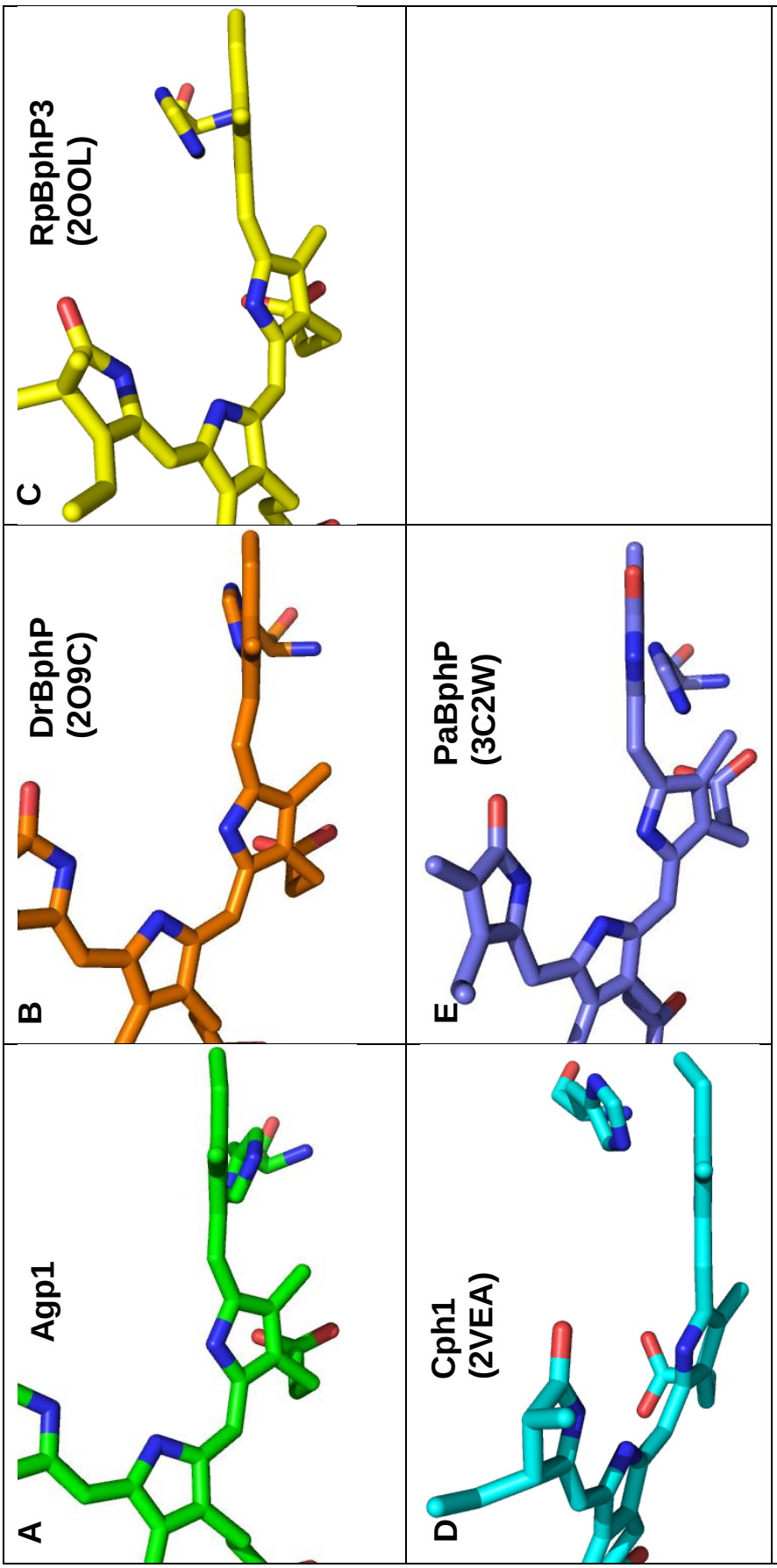


Figure 3.38 Conformation of the conserved His residue relative to the D-ring of the bilin moiety. A) Agp1, B) DrBphP (2O9C), C) RpBphP3 (2OOL), D) Cph1 (2VEA), E) PaBphP (3C2W). His 280 of Agp1 lies on the β-facial side of the D-ring unlike other phytochromes in the Pr-form, but similar to PaBphP in the Pfr-form. The relative orientation of the imidazole ring with respect to the D-ring in Agp1 is unlike all other phytochromes. Relative point of view is not constant, in order to clarify the His orientation relative to the D-ring.

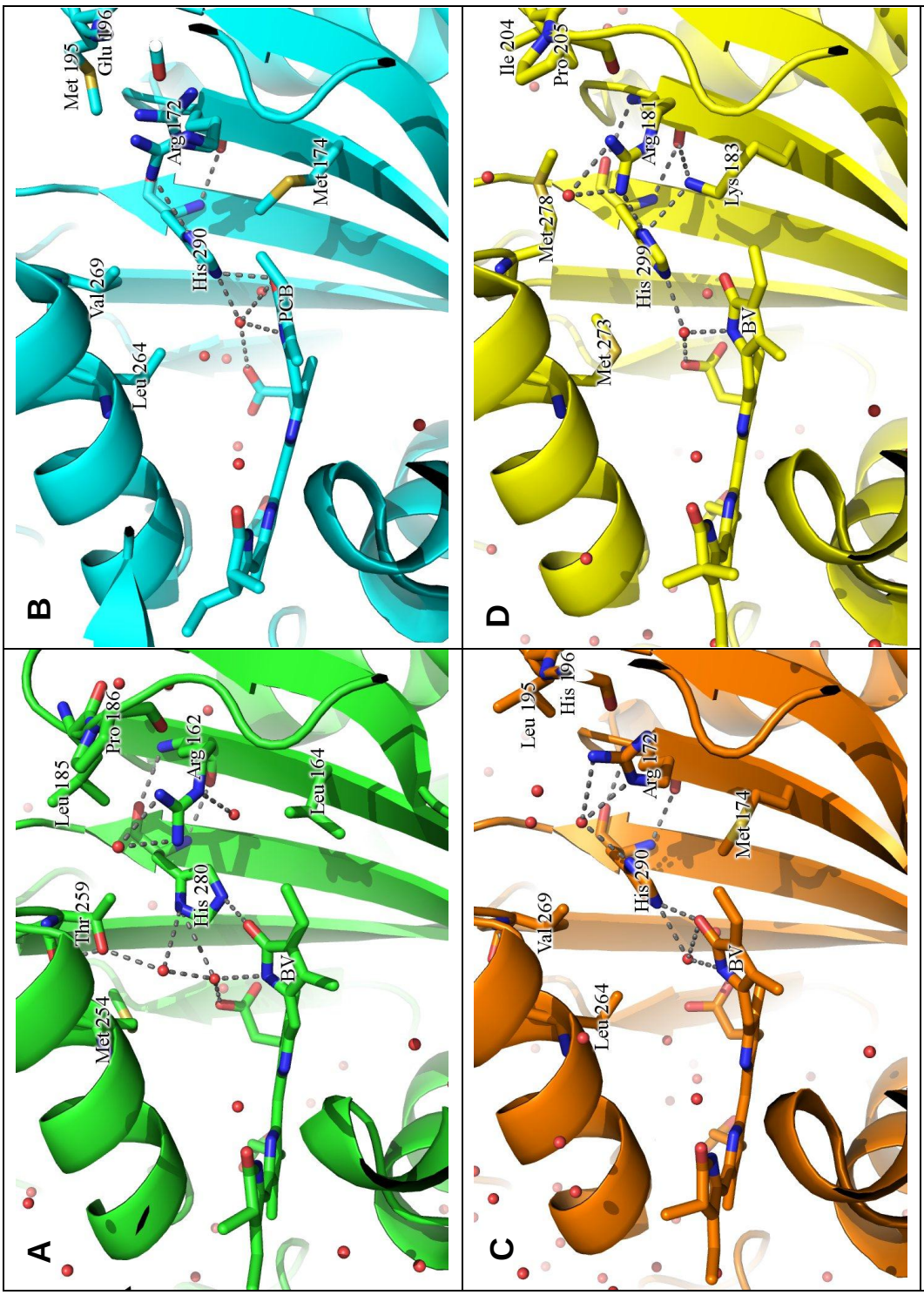


Figure 3.39 Differences between Agp1 and other phytochromes in the Pr-form in the vicinity of the D-ring and the C-ring propionate side chain. A) Agp1, B) Cph1 (2VEA), C) DrBphP (2O9C)Green, D) RpBphP3 (2OOL).

3.10.8 Solvent shield surrounding the bilin moiety

Covalent bond between Cys 20 and BV in Agp1 appears similar to that of DrBphP, whose mode of binding has been used as the model (Wagner 2007). In Cph1, Leu 15 corresponds to the chromophore-binding Cys in the N-terminus of bacteriophytochromes, whereas mostly hydrophobic residues occupy the corresponding position of the chromophore-binding Cys in the GAF domain of Cph1 in bacteriophytochromes (Agp1 Val 249, DrBphP Met 259, PaBphP Ile 246, RpBphP3 Thr 268). In Agp1, Ile 459 – Pro 461 of the *Tongue* (Cph1: Leu 469 – Pro 471, Essen et al., 2008) forms a hydrophobic shield around the thioether linkage, but Tyr 458 of Cph1, another component of the solvent shield, is not well conserved in bacteriophytochromes. Agp1 Thr 452 occupies the equivalent location of Cph1 Tyr 458 with the side chain protruding in a similar direction. In Agp1, Thr 452 is in proximity of the thioether linkage between the BV and Cys 20, but the distance between Thr 452 and the thioether linkage is clearly more than if the position of Thr 452 is occupied by a larger Tyr residue. Cph1 Tyr 458 is close to the A-ring of the PCB that its van der Waals force is recognised by the default setting of the Ligplot (Fig. 3.41), but the thioether linkage in Cph1 protrudes away from Tyr 458, unlike in Agp1. Presumably the acquisition of a residue with a bulky side chain to act as a solvent shield in this position is an evolutionarily adaptation associated with the positional change of the chromophore-binding Cys from the N-terminus to the GAF domain.

Interestingly, in DrBphP structure, the space corresponding to the solvent shield (Agp1 Ile 459 – Pro 461) is partially occupied the N-terminal region (Fig. 3.40). The structure of DrBphP lacks the PHY domain, therefore its structure may be an artefact. Sequence alignment however, supports the notion for the N-terminal region to act as the solvent shield in DrBphP. Residues equivalent for the solvent shield of the *Tongue* are present in DrBphP (Leu 463 – Pro 465), but there is a deletion of the sequence unique to DrBphP in the vicinity, therefore the tip of the *Tongue* of DrBphP will not shield the BV to the same degree as in Agp1 or Cph1, making it very likely that the same N-terminal conformation as observed in the DrBphP structure is also adopted in the full-length structure since it is required for the solvent shield by compensating for the shorter *Tongue* of the PHY domain in DrBphP.

The residues equivalent to the above-mentioned solvent shield from the *Tongue* is present in PaBphP (Leu 454 – Pro 456), but they are not spatially conserved in the WT

structure. The lack of the solvent shield from the *Tongue* is partially compensated by Leu 9 in PaBphP. This is likely to be the result of PaBphP assuming the Pfr-form, however.

Cph1 does not have a Pro residue immediately prior to Cys 259 in the GAF domain. The Pro residue is conserved among all structurally determined bacteriophytochromes and *SyB*-Cph1. This Pro can be interpreted to be the ancestral trait functioning to provide rigidity to the N-terminal end of the α8 of Agp1 and the corresponding region in other phytochromes, but may have become functionally redundant for Cph1 because formation of covalent bond at the GAF domain limits the conformational freedom. Therefore the lack of the Pro residue in Cph1 may be the result of stochastic silent mutation undirected by evolutionarily pressure.

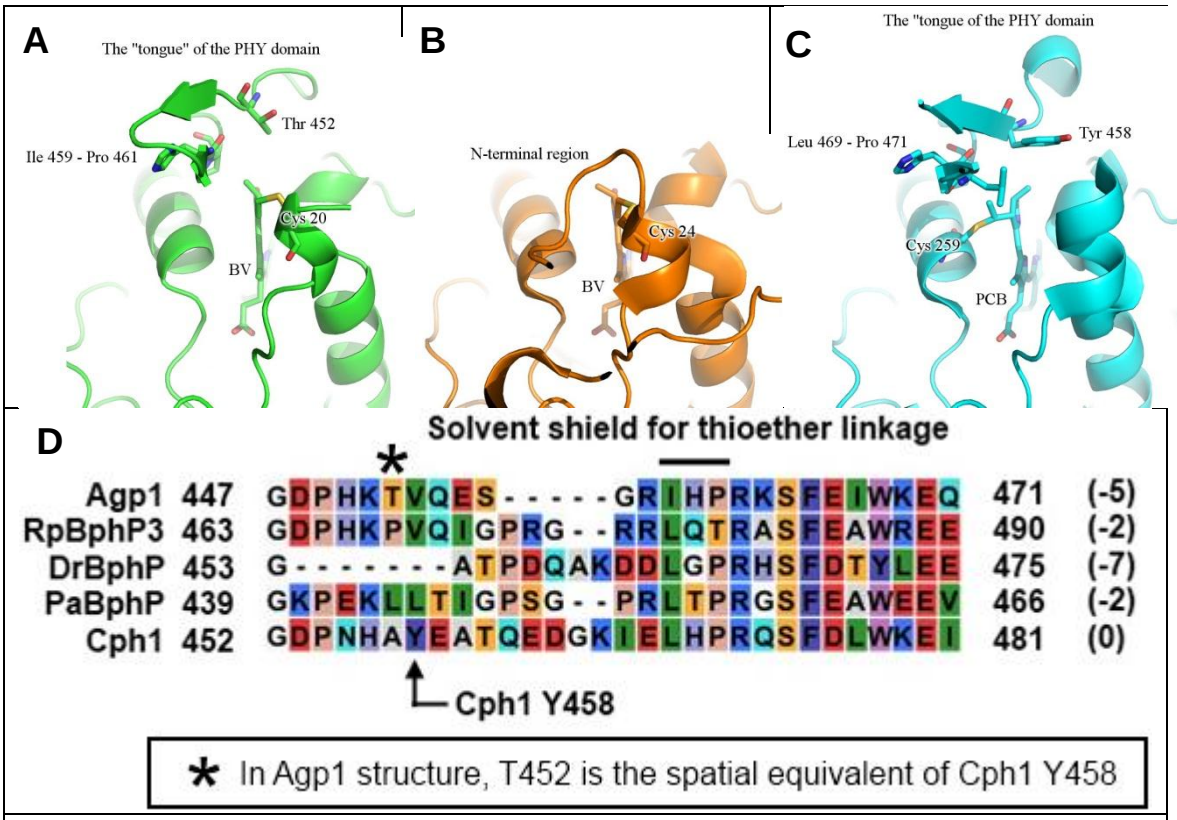


Figure 3.40 Solvent shields around the thioether linkage. Cartoon and stick model of A) Agp1, B) DrBphP, and C) Cph1. D) Sequence alignment of a part of the *Tongue*. The numbers in parentheses indicate the difference in the number of residues of this region relative to Cph1. Cph1 Tyr 458, a bulky solvent shield is only found in Cph1. This is regarded to be an evolutionarily development associated with the change in the chromophore-binding Cys residue from the N-terminus to the GAF domain. The N-terminal region of DrBphP partially occupies the space corresponding to the solvent shield of the PHY domain of other phytochromes. This is likely to compensate for a relatively shorter tip of the Tongue, and the N-terminal conformation is likely to be the same in the full-length structure.

3.10.9 Propionate side chains

Some molecular interactions in the vicinity of propionate side chains of C- and D-rings were found to differ between Cph1 and DrBphP, and this is thought to reflect the differences between bacterial and cyanobacterial phytochrome (Essen et al., 2008), which may give rise to different types of photochemistry between them (Rockwell et al., 2009).Consistent with this idea, in terms of molecular interactions, Agp1 resembles DrBphP more closely than Cph1.

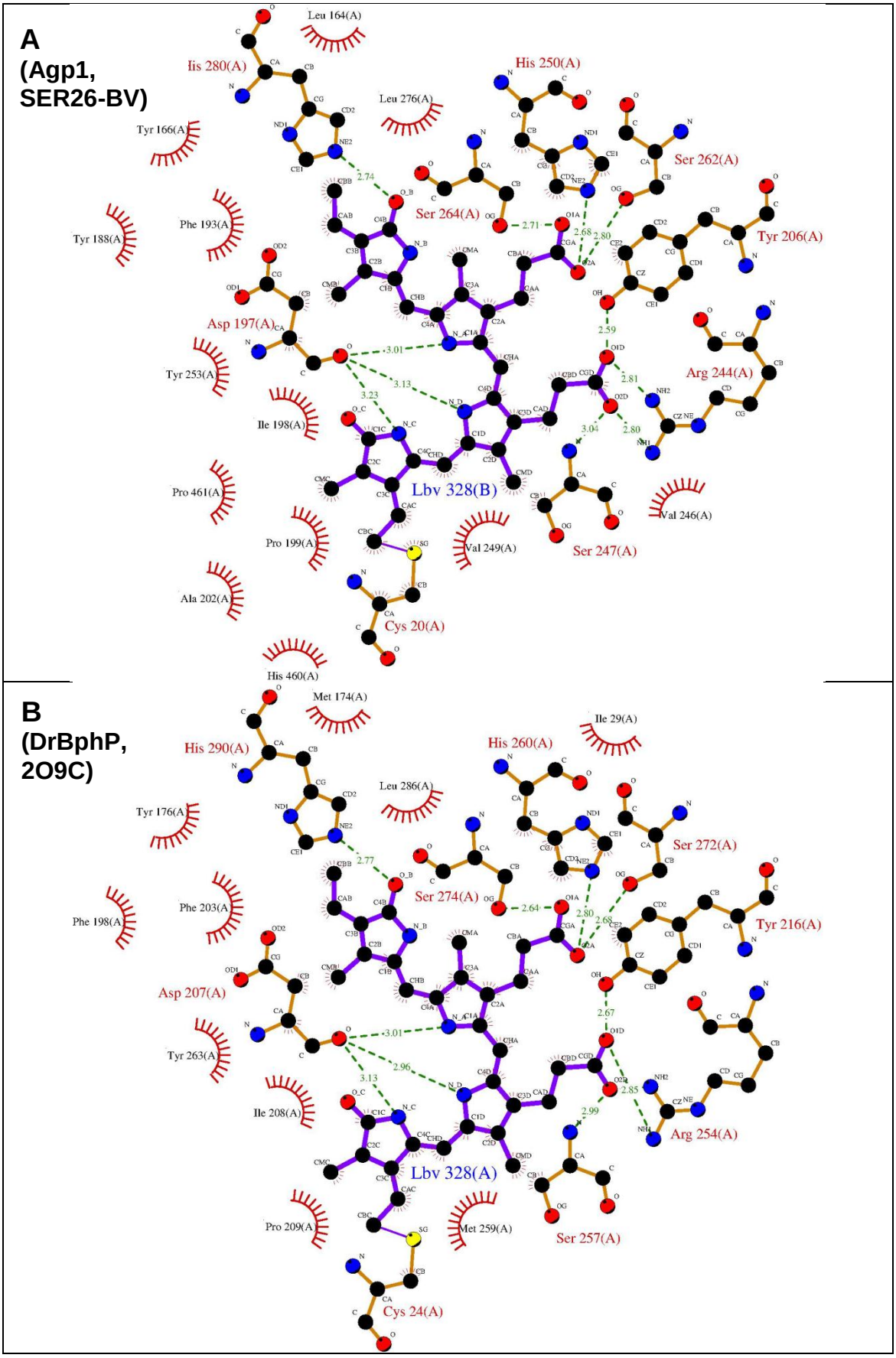
Tyr 206, Arg 244, and Ser 247 interact with the B-ring propionate of BV in Agp1. The salt bridge between the B-ring propionate and Arg 244 is conserved in all canonical phytochromes in the Pr-form. The interaction between the Tyr and the Ser with the B-ring propionate is conserved among bacteriophytochromes in the Pr-form (DrBphP and RpBphP3). In Cph1, the equivalent interaction between the B-ring propionate and the N-atom of Ser 247 in Agp1 is realised by the N-atom of Tyr 257 in Cph1, but the equivalent polar interaction between the OH-group of Tyr 206 in Agp1 is not possible in Cph1 due to the absence of the hydroxyl group in Phe 216. An Ile residue which forms a hydrophobic walling around the A- and the B-ring of the bilin in other phytochromes is also present and is conformationally conserved in Agp1. The shortest distance between two moieties are longer in Agp1, however, and the hydrophobic interaction by the Ile residue is not recognised by Ligplot with the default parameters (Agp1: Ile 25 CG2 – BV CMB 4.1Å, DrBphP (2O9C): Ile 29 – BV CMD 3.9Å, Cph1 (2VEA): Ile 20 – PCB CBD 3.6Å) (Fig. 3.41).

The propionate side chain of the C-ring of the BV in Agp1 interacts with His 250, Ser 262, and Ser 264, all of which are invariably conserved among structurally defined bacteriophytochromes. The interaction between Ser 264 and the C-ring propionate is conserved among bacterial phytochromes in the Pr-form. The importance of this residue for maintaining the Pr-form has been demonstrated by the Ser261Ala mutant in PaBphP which shows no detectable formation of the Pr-form upon illumination (Yang et al., 2009). As similar interaction is present in Cph1 involving Thr 274, but its larger side chain relative to Ser induces the carboxyl group of the C-ring propionate to be rotated by about 90° relative to that in bacteriophytochromes. The conformation of the C-ring propionate side chain of PCB in Cph1 does not allow the formation of a hydrogen bond equivalent to that between His 250 and the C-ring propionate in Agp1.

A hydrogen bond is present between C-ring propionate and Ser 275 in PaBphP, but the position of the latter is occupied by Ala in classical phytochromes (Agp1, DrBphP, and Cph1) preventing such an interaction (Yang et al., 2009).

Exceptionally, the conformation of Ser 262 of Agp1 is more similar to the equivalent Ser 272 in Cph1 rather than DrBphP, but the interaction between the Ser residue and the C-ring propionate is similar between Agp1 and DrBphP as regards the direct hydrogen bond between the two moieties, whereas the interaction between the C-ring propionate and Ser 272 in Cph1 is mediated by a water molecule (Fig. 3.42).

These subtle differences in phytochromes between species could be the result of the selective pressure imposed on the host organisms which resulted in fine-tuning of the photochemical activities. Thus, the differences in conformation and the subsequently the hydrogen bonding network may underpin the differences in photochemical properties between bacterial and cyanobacterial phytochromes. One such difference is the ability of Cph1, a cyanobacterial phytochrome, to maintain the photoconverted Pfr-form without the dark-reversion, which is unlike bacterial phytochromes.


A
(Agp1,
SER26-BV)
Leu 164(A)
Iis 280(A)
His 250(A)
Leu 276(A)
Tyr 166(A)
Ser 262(A)
Ser 264(A)
Phe 193(A)
Tyr 188(A)
Tyr 206(A)
Asp 197(A)
Arg 244(A)
Tyr 253(A)
Ile 198(A)
Lbv 328(B)
Pro 461(A)
Val 246(A)
Pro 199(A)
Val 249(A)
Ser 247(A)
Ala 202(A)
Cys 20(A)
His 460(A)
B
(DrBphP,
2O9C)
Met 174(A)
Ile 29(A)
His 290(A)
His 260(A)
Leu 286(A)
Tyr 176(A)
Ser 272(A)
Ser 274(A)
Phe 203(A)
Phe 198(A)
Tyr 216(A)
Asp 207(A)
Tyr 263(A)
Ile 208(A)
Arg 254(A)
Lbv 328(A)
Pro 209(A)
Met 259(A)
Ser 257(A)
Cys 24(A)

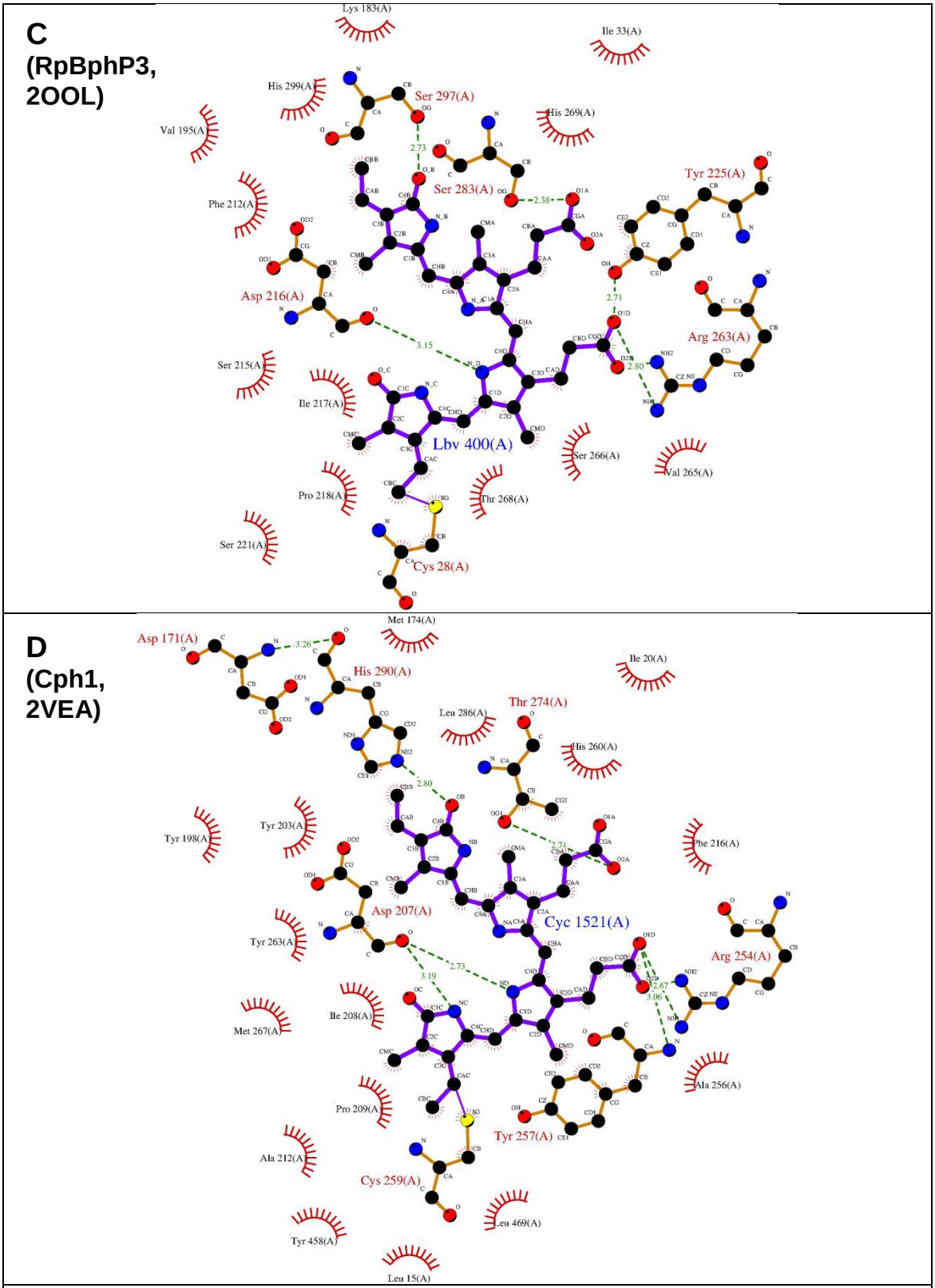


Figure 3.41 Molecular interactions of the bilin moieties in phytochrome structures. Water molecules are omitted. Diagram produced by Ligplot (Wallace et al., 1995). A) Agp1 SER26-BV, B) DrBphP (2O9C), C) RpBphP3 (2OOL), D) Cph1 (2VEA). Orientation of the bilin moiety and the relative positions of the equivalent residues have been adjusted to Agp1 as the reference. Positions of the conserved residues that are not found in Agp1 have also been adjusted to be consistent with one another in terms of relative position to the bilin moiety on the diagram.

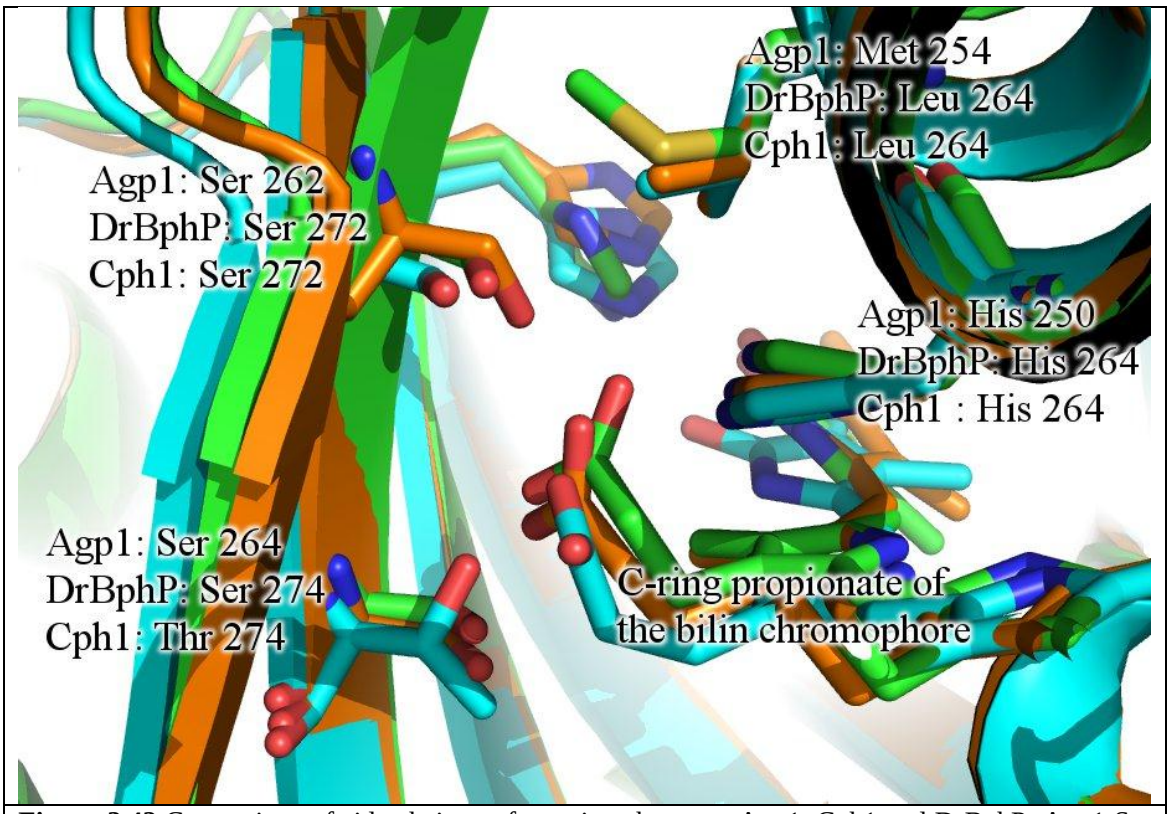


Figure 3.42 Comparison of side chain conformations between Agp1, Cph1 and DrBphP. Agp1 Ser 264 is conserved with DrBphP, but conformational conservation is exhibited at Agp1 Ser262 is with Cph1. Met 254 imposes less van der Waals force than Leu, which ultimately contribute to C-ring being shifted towards it compared to the others.

3.10.10 Structural differences between the Pr and Pfr-forms at the molecular level

Structural comparison between the PCM fragments of Agp1 and Cph1 reveals similarities between them. This is significant, because the observation is despite a relatively long phylogenetic distance between Agp1, a bacteriophytochrome and Cph1, a cyanobacterial phytochrome (Lamparter, 2004). Therefore the similar conformation between them is deemed the true representative of the Pr-form as originally suggested by Essen and co-workers (Essen et al., 2008).

The same argument could be made for the structures of the bathy-phytochromes, PaBphP and RpBphP1, the latter of which was determined recently (Bellini and Papiz, 2012b, Yang et al., 2008). The similarities between PaBphP and RpBphP1 include the global conformation and the secondary structure of the *Tongue* of the PHY domain. Yet the similarities are despite the lack of conservation of the residues that stabilise the Pfr-from, for example the equivalent interactions to those involving Arg 453 and Gln 188 are not present in RpBphP1 structure (Bellini and Papiz, 2012b). The similarities exhibited by two independent phylogenetically distant phytochromes may indicate that these structures represent the Pfr-form for all phytochromes.

Many structural differences are evident form comparing the structures of the phytochromes in the Pr- and the Pfr-forms, and the differences between the two forms may be attributed to differences at the primary structure level. Interestingly, however, there are also examples of strongly conserved residues participating in a key interaction in one form but not in the other form.

The interaction that stems out of the conserved "PAxDIP" motif differs between Agp1/Cph1 in the Pr-form and PaBphP in the Pfr-form. In classical phytochromes in the Pr-form, the Asp (Agp1 Asp 197) within the PAxDIP motif forms a hydrogen bond with a Tyr (Agp1 Tyr 253) and a salt bridge with an Arg (Agp1 Arg 462), the latter of which is a part of conserved PRxSF motif (Essen et al., 2008). The Arg in turn forms a hydrogen bond with the oxygen atom of the second Gly of the conserved $W^{G}/_{A}G$ motif (Agp1 Gly 446), but an additional interaction between Arg 472 and Ala 457 in Cph1 is absent in Agp1, due to the conformational difference at the *Tongue* in which the latter residue is located. Interestingly, none of these interactions occurs in PaBphP, despite conservation of the PAxDIP motif, PRxSF motif, the equivalent of Agp1 Tyr 253 (PaBphP Tyr 250) and the $W^{G}/_{A}G$ motif. The PaBphP Asp 194 in the PAxDIP motif interacts with Arg 453 and Ser 459, both of which are widely conserved (Yang et al.,

2008). The equivalent interactions involving the homologous Arg 458 and Ser 464 in Agp1 are absent. There are countless instances where strongly conserved residues are in different conformations. Interdomain polar interactions between the GAF and the PHY domains are shown overlaid onto a sequence alignment in Fig. 3.43, and it demonstrates the fact that some identical amino acids exhibit context dependent conformation.

A high degree of conservation of the "$W^{G}/_{A}G$" motif (Agp1 445-447) in the *Tongue* and its role in establishing an unusual conformation in Cph1 has been reported (Essen 2008). In Agp1, the hydrogen bonds involving the $W^{G}/_{A}G$ motif is conserved with Cph1. Inspection of the main chain torsion angles of the equivalent residues in Agp1, Cph1 (2VEA), PaBphP (3C2W) and RpBphP1 (4EHO) suggests that phytochromes in the Pr-form (Agp1 and Cph1) may not require a Gly residues for rigidly maintaining the conformations found in the crystal structures (Fig. 3.44). Yet interestingly the $W^{G}/_{A}G$ motif is widely conserved among phytochromes, indicating the presence of selective pressure to maintain these residues. The torsion angles of this region in PaBphP (Pfr-form), however, indicate that a glycine is required for its conformation found in the crystal structure. Perhaps this observation supports a hypothesis for phytochromes to undergo dynamic conformational change during photoconversion from the Pr to the Pfr-form that is described later. Examples where the presence of a glycine facilitates large conformational change by acting as a "hinge" have been reported, for example DNA-binding HapR protein (Dongre et al., 2011). An analysis with RpBphP1 shows that the dihedral angles of the $W^{G}/_{A}G$ motif are within the allowed region of the Ramachandran plot for the general amino acids. A wide range of dihedral angles (Table 3.13), together with disorder found in some chains of the crystal structures of PaBphP and RpBphP1 in the $W^{G}/_{A}G$ motif, is consistent with an idea of a dynamic motion involving the $W^{G}/_{A}G$ motif.

In classical phytochromes the conserved Trp residue (Agp1 Trp 445, Cph1 Trp 450) of the $W^{G}/_{A}G$ motif is situated close to the bilin pocket, and undergoes a hydrophobic interaction with the first proline of the "PAxDIP" motif (Agp1 Pro 194, Cph1 Pro 204), as well as forming a polar interaction with a Glu residue (Agp1 Glu 470, Cph1 Glu 480) via their side chains. An interesting case of spatial compensation is found in PaBphP, where the space equivalent to Agp1 Trp 445 relative to the GAF domain is occupied by Trp 463 (Agp1 Trp 468), and this is despite the equivalent of Agp1 Trp 445 is present in PaBphP in the form of Trp 437. The equivalent of Agp1 Glu 470 is also present in PaBphP as Glu 465, but neither is the equivalent interaction between Agp1 Trp 445 –

Glu 470 present in PaBphP nor does Glu 465 form any polar interaction with its side chain.

It is noteworthy that some residues in regions with completely different conformations are widely conserved in phytochromes. If a residue makes an interaction via a main chain atom, the selective pressure to conserve its side chain will not be as strong as a case where an interaction occurs via its side chain. For example, a case of conserved interaction via main chain atoms despite dissimilar amino acid content is shown in Section 6.5.2. Observations that several identical amino acid residues are found to assume different conformations in the Pr- and Pfr-forms suggest the importance in maintaining those residues, especially if such a residue participate in an interaction in one form, but the corresponding interaction is absent in the other form. The reason behind a high degree of amino acid conservation in context dependent conformation may be because the PHY domain undergoes conformational change between that found in Pr (Agp1 and Cph1) and Pfr (PaBphP and RpBphP1) forms, thereby the switching of conformation poses selective pressure even on a seemingly unimportant residue in one form that makes a key interaction in the opposite form.

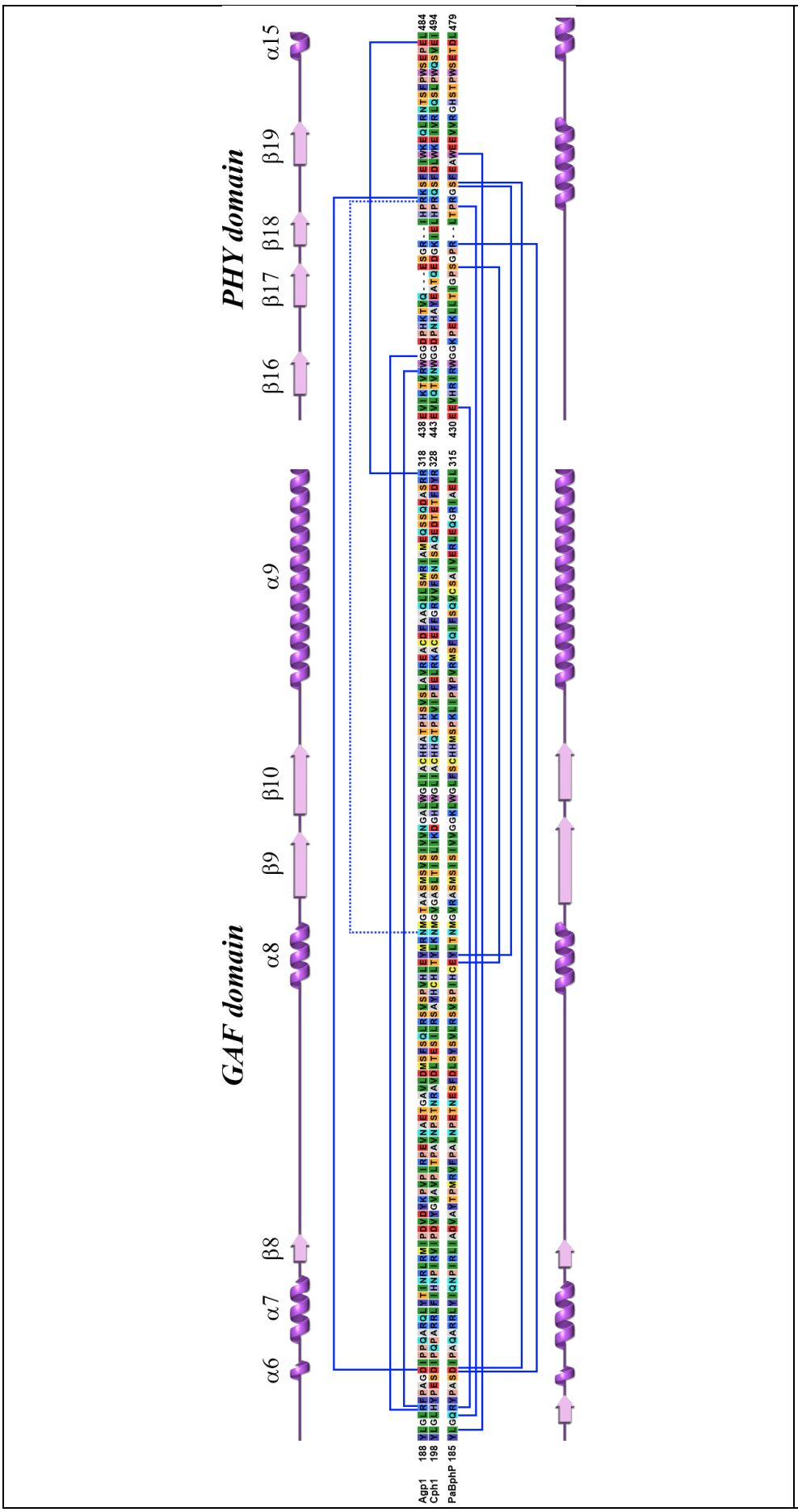
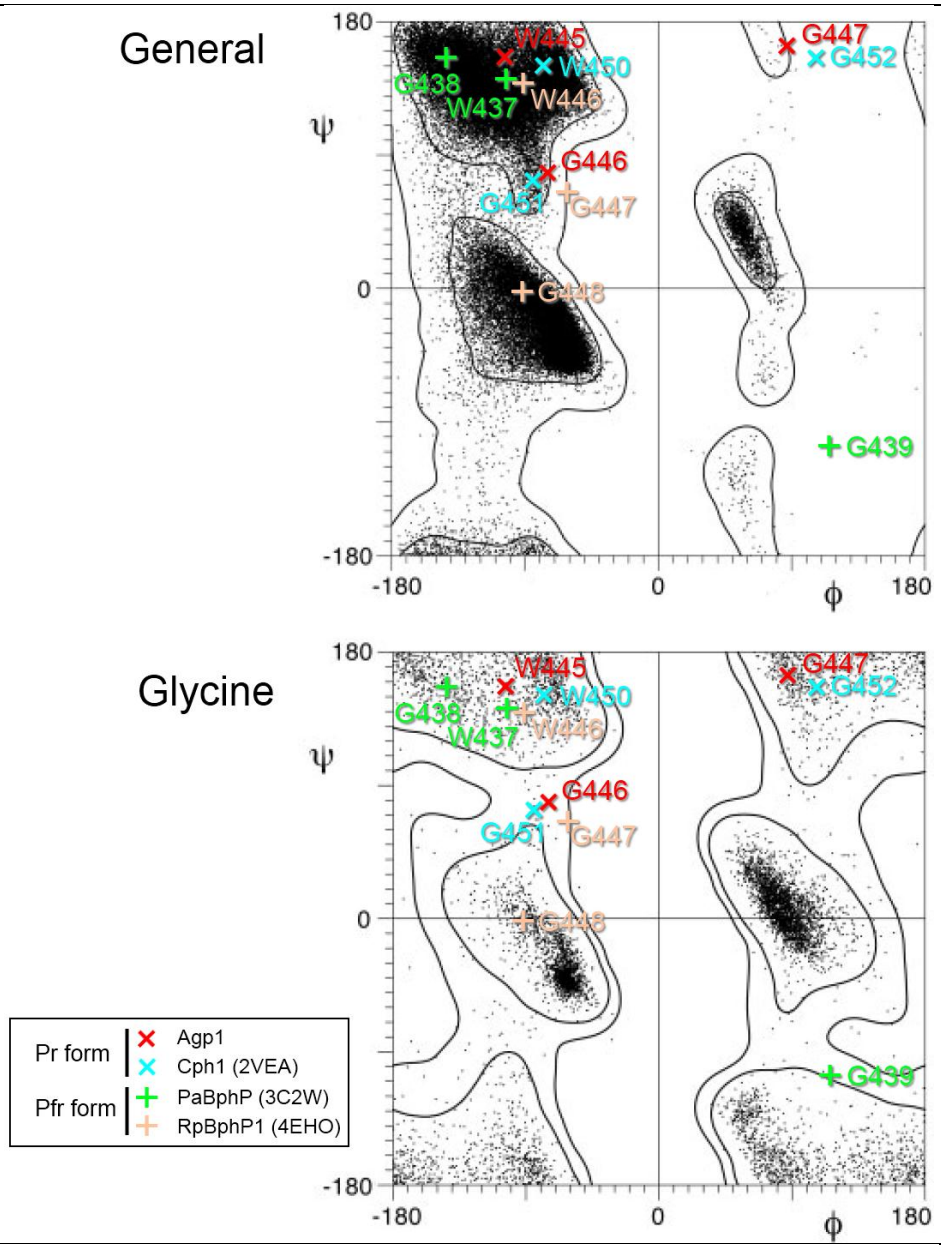


Figure 3.43 Different patterns of polar interaction between phytochromes in the Pr- and the Pfr-forms. Inter-domain interactions between the GAF and the PHY domains as identified by Ligplot are overlaid onto aligned sequences of Agp1, Cph1, and PaBphP. An interaction unique to Agp1 is displayed in a dotted line. The secondary structures of Agp1 (SER26-BV) and PaBphP (3C2W) are also shown. No inter-domain interaction is identical between the Pr- and the Pfr-forms, despite the participating residues themselves are often conserved.

A



B

			W	G	G
Pr-form	**Agp1**	Phi	-104.02	-74.13	87.64
		Psi	157.77	79.49	163.75
	Cph1 (2VEA)	Phi	-103.96	-86.21	107.72
		Psi	150.71	73.60	167.48
Pfr-form	**PaBphP (3C2W)**	Phi	-103.96	-144.07	116.94
		Psi	142.49	157.20	-107.01
	RpBphP1 (4EHO)	Phi	-90.27	-62.34	-92.37
		Psi	139.88	65.74	-2.22

Figure 3.44 Torsion angles of the $W^{G}/_{A}G$ motif in Agp1, Cph1 (2VEA), PaBphP (3C2W) and RpBphP1 (4EHO). The "A" chain was used where multiple chains are available. A) The Phi and Psi angles are overlaid onto the Ramachandran plots for general amino acids (non-glycine and non-proline residues) and glycines. The torsion angles of the $W^{G}/_{A}G$ motif of Agp1 and Cph1 are either within or very close to the allowed regions for the general amino acids, which may suggest that glycines may not be required to maintain the conformations found in the crystal structures. The torsion angles of Gly 439 of PaBphP lie far from the allowed regions for general amino acids. Ramachandran plot diagrams adapted form Lovell et al. (2003). B) Tabulated torsion angles of the $W^{G}/_{A}G$ motif in three phytochromes. Values are in angles (°).

	Chain		W	G	G
PaBphP (3C2W)	A	Phi	-103.96	-144.07	116.94
		Psi	142.49	157.20	-107.01
	B	Phi	N/A	N/A	N/A
		Psi	N/A	N/A	N/A
	C	Phi	-92.43	-166.96	92.02
		Psi	151.19	-170.33	-160.31
	D	Phi	-92.20	-145.22	N/A
		Psi	117.49	N/A	N/A
	E	Phi	-121.15	-172.80	150.21
		Psi	159.80	162.19	N/A
	F	Phi	N/A	N/A	N/A
		Psi	N/A	N/A	N/A
	G	Phi	-98.69	-71.31	-74.31
		Psi	101.76	109.41	57.31
	H	Phi	-82.39	-52.69	-153.83
		Psi	20.02	117.05	175.62
RpBphP1 (4EHO)	A	Phi	-90.27	-62.34	-92.37
		Psi	139.88	65.74	-2.22
	B	Phi	54.90	-58.05	-80.48
		Psi	52.96	116.73	59.17
	C	Phi	-138.80	-179.85	-172.63
		Psi	-116.18	132.77	127.13
	D	Phi	N/A	N/A	N/A
		Psi	N/A	N/A	N/A

Table 3.13 Torsion angles of the WG/AG motif all the chains of the crystal structures of PaBphP (3C2W) and RpBphP1 (4EHO). Dihedral angles of the corresponding residues range wide, and the $W^{G}/_{A}G$ motif was disordered in some chains. Altogether, an inherent flexibility of the region appears to be conserved in phytochromes in the Pfr-form. Values are in angles (°).

3.11 Discussion and future works

3.11.1 M20 holoprotein

A shorter fragment of Agp1, M20 either with or without the first nine N-terminal residues, failed to produce suitable crystals for structural determination reproducibly, despite proteins produced from these constructs were subjected to the most intensive screening in the present study. Apart from the possible failure to find a suitable crystallisation condition, the failure to crystallise M20(s)-BV may be attributed to its physical properties such as detailed below.

The amino acid composition of the surface of M20 may be thermodynamically unfavourable for crystallisation, either because of the lack of suitable crystal epitopes or a higher proportion of amino acid residues with high conformational entropy than in M15. This hypothesis was tested by calculating the proportion of solvent accessible lysine and glutamate, the primary target residues for the SER strategy, on the surface of the monomeric models of SER26, M15 (WT), and M20 (WT). Structural models of the latter two were generated based on the SER26-BV structure using PyMOL, and analysed using Areaimol. The proportions of solvent accessible Lys and Glu residues on the surface of SER26, M15, and M20 obtained by this method are 12.8%, 17.6%, and 15.0%, respectively. M20 has a lower proportion of the SER target residues on the surface than readily-crystallisable M15 protein, therefore the recalcitrance of M20 to crystallise cannot be solely accounted for by the surface entropy.

An alternative hypothesis for rationalising the difficulty to crystallise M20 involves the oligomerisation propensity of M20. The oligomerisation of proteins can be positively correlated with the success of crystallisation trials according to a statistical study from a structural genomic project (Price II et al., 2008). The M20 protein consisting only of the CBD is less likely to form a dimer compared to M15 protein consisting of the PCM, because the former has less dimer interface area due to the lack of the PHY domain (Section 6.2.1).

A recent structural study on RpBphP2 CBD further supports the notion of dimerisation to be important for successful crystallogenesis. The study indicated a single arginine to play a key role for the formation of a dimer by the virtue of its long side chain (Bellini and Papiz, 2012a).

The fact that the M20 dimer exists as a lower proportion of the total sample is attributed to two factors. Firstly, M20 has lower dimer interface area compared to M15, and secondly a Thr (position 131) occupies the position corresponding to Arg 150 in RpBphP3. Crystallisation of PaBphP as a parallel dimer, despite the presence of Gln residue in the corresponding position, was attributed to the presence of the PHY domain which provides an additional dimer interface (Bellini and Papiz, 2012a), and the same may be assumed for the crystallisation of WT Agp1 M15(s) as a parallel dimer.

3.11.2 M15 holoprotein

Despite the accurate phase information supplied by molecular replacement with SER26-BV coordinates, the structure of M15-BV could not be refined satisfactorily. The possibility of pseudosymmetry was addressed by processing the data with lower symmetry space groups, while using the twin refinement option of Refmac5 where applicable did not result in any appreciative improvement. Moreover, Phenix.refine, an alternative refinement program was interchangeably used with Refmac, but the exact source(s) of the problem could not be identified, and the final R/R_{free} values were 0.2773/0.3586 with undefined electron density in some areas, including the BV moiety.

3.11 3 SER26-BV

Application of the SER approach effectively addressed the “pathological” case by changing the crystal form, and led to the determination of Agp1 structure, albeit as a biologically irrelevant anti-parallel dimer. The detailed effect of SER mutations are discussed in Section 5.5.

3.11.4 Structural comparison of Pr and Pfr-forms of phytochromes

Structural observations were carried out in an attempt to relate structural features to the underlying variables, and in particular to draw out key differences between Pr and Pfr-forms. It must be made clear prior to any conclusions, however, that the structure of PaBphP in the Pfr-form (ground state for bathy phytochrome) may not represent the Pfr-form of classical phytochromes (photoconverted state). An additional question may be posed over whether the PaBphP structure is representative for the entire bathy-phytochrome family, based on a phylogenetic study of bathy-phytochromes from bacteria from *Rhizobiales* order (Rottwinkel et al., 2010). The Rottwinkel study presents a hypothesis for several bathy-phytochromes (PaBaphP, type 1; *Rhizobiales* bathy

phytochromes, type 2) to have separate origins. In that case, the bathochromic ground state in these bathy-phytochromes must be the result of convergent evolution.

Sequence comparisons between Agp1, Cph1, and PaBphP reveal differences in the primary structure between classical and bathy-phytochromes at the GAF and the PHY domains, e.g. Q188 and Ser 275, which are unique to PaBphP (Yang et al., 2009), or Agp1 Pro 376 that is unique to classical phytochromes. They have become presumably adapted for stabilising the bilins in different conformations. On the other hand, many conserved amino acid residues adopt different conformations in classical and bathy phytochromes. This kind of context-dependent conformation begs a question: what is the identity of the determinant factor for the protein conformation, particularly at the *Tongue* of the PHY domain? Based on the crystal structures of the PaBphP Q188L mutant in mixed Pfr/Pr-forms, a gross conformational difference found between phytochromes in the Pr-form (Cph1, DrBphP) and the Pfr-form (PaBphP) was deemed unlikely to reflect the molecular trajectories of the photoconversion (Yang et al., 2009). It is possible, however, that Q188L mutation in PaBphP affected not only the rate of dark reversion (Pr to Pfr), but also some point, in the chain of events that is simply referred to as "photoconversion", is perturbed. Therefore the structure of PaBphP Q188L mutant in the spectrophotometric Pr-form may not exhibit the natural conformation of the Pr-form. For example, *A. thaliana* PhyB mutants with normal spectrophotometric property but deficient in signal transduction have been reported (Nagatani, 2010). To date, no evidence has been presented to verify the biochemical properties of the PaBphP Q188L mutant in the spectrophotometric Pr-form to represent that of the WT, for example by electron microscopy (global conformation in solution) or kinase assay (signal transduction).

As an alternative picture of photoconversion, a dynamic conformational change, where conformations of the conserved amino acids switch from those observed in the Pr-form (Agp1 and Cph1) and the Pfr-form (PaBphP), is possible. According to this model, the PaBphP structure truly represents the Pfr conformation of the classical phytochromes. The residues that make key interactions either in classical phytochromes in the Pr-form or bathy-phytochromes in the Pfr-form are highly conserved in all phytochromes, despite their conformations being different. Also, the presence of a highly conserved $W^{G}/_{A}G$ motif may suggest mobility of the *Tongue*. This raises another question of whether the PHY domain of a classical phytochrome can adopt the conformation similar to that of bathy-phytochromes, and vice versa. It will not be entirely impossible, though

thermodynamically unfavourable compared to the native ground stable form, therefore even if an alternative conformation should arise (the Pfr-form for the classical, and the Pr-form for the bathy phytochromes), the conformation will soon revert back to the thermodynamically stable form.

With regards to what determines the stable conformation, I propose that not only the limited numbers of non-conserved residues unique to each type (classical and bathy), but also numerous other residues at the GAF-PHY interface by both domains, to synergistically determine the most thermodynamically favourable interaction. Such an idea is consistent with the observations that mutational studies explored thus far, to modify the spectrophotochromic property of phytochromes from one type to the other (e.g. bathy to classical or unusual to classical and vice versa), have never resulted in full conversion, in spite of attempts involving either single mutations (Yang et al., 2008, Yang et al., 2009), quadruple mutations in one domain (Yang et al., 2007) or even a whole domain swap (Zienicke et al., 2011). Mutation(s) in a single domain or single domain swap will shift the equilibrium of the favoured conformation of the phytochrome, but would be insufficient for the full conversion. The thought that dynamic conformational change might be associated with photoconversion is consistent with global conformational differences between the Pr- and the Pfr-forms, as demonstrated by analyses such as SEC (Esteban et al., 2005, Inomata et al., 2005, Lamparter et al., 2001) and limited proteolysis (Noack et al., 2007).

The photochemical properties of phytochrome fragments published thus far leads to the consensus that the CBD is sufficient for bilin lyase activity as well as formation of the Pr-form, but insufficient for fully photoactive holoprotein (Noack et al., 2007, Wagner et al., 2005, Yang et al., 2008, Yang et al., 2007). There are exceptions to this, however. *SyB*-Cph1 can form the Pfr-form only with the GAF domain, while Agp2 only loosely accords to consensus (Cornilescu et al., 2008, Karniol et al., 2005). Unique aspects of *SyB*-Cph1, a member of the "PAS-less phytochrome" group with a non-canonical domain architecture may be the result of divergent evolution, so that the roles allotted to each domain have changed. The formation of the Pfr-form requires the PHY domain in both classical and bathy-phytochromes with the canonical domain organisation, and indeed the CBD of PaBphP was reported to be in the Pr-form (Yang et al., 2008). The majority of the mutants of PaBphP eventually revert to the Pfr-form in darkness, except for a D194A mutant which constitutively adopt the Pr-form and unable to undergo photoconversion (Yang et al., 2008). The latter case is exceptional because a mutation

of the conserved Asp has been shown to affect the formation of the Pfr-form in other phytochromes (Hahn et al., 2006, Wagner et al., 2008, Yang et al., 2007). Overall, the ground stable form of a phytochrome and the associated stereochemistry of the bilin are ultimately determined by the protein conformation, which itself is determined by the synergistic interaction between the GAF and PHY domains.

Chapter 4: Structural characterisation of Agp1 in the Pfr-form

4.1 Overview

Information regarding the molecular trajectories during a phytochrome photoconversion event will provide insights into the mechanism of photoconversion and the signal transduction. *SyB*-Cph1 and PaBphP in their photoconverted state have been structurally characterised (Ulijasz et al., 2010, Yang et al., 2009). However, the structures of *SyB*-Cph1 and PaBphP in their photoconverted states are not ideal for deducing the mechanism of signal transduction and global conformational change, respectively (Section 1.3.4). It is therefore desirable to structurally characterise the PCM of a single phytochrome in both Pr and Pfr-forms independently. Any structural differences between two structures are accounted for solely by the difference in their states. To this end, a crystallographic study was undertaken to probe the structure of Agp1 in the Pfr-form.

Crystallisation of Agp1 in the Pfr-form is difficult, however. Bacterial phytochromes in the Pfr-form are relatively unstable because they undergo dark-reversion from the photoconverted state into the ground state (Section 1.3.3). Full-length Agp1-BV undergoes dark reversion with initial rate of 3.4×10^{-4} s^{-1} which means that in the Pfr-form completely disappears on an hour time scale (Lamparter et al., 2002). The dark reversion renders the Pfr-form of Agp1 a difficult target for crystallisation, because the time scale of the dark reversion is faster than the typical time scale required for crystallisation which is in the order of days to weeks. Cph1, a classical cyanobacterial phytochrome, is known to remain stable in its photoconverted state (Pfr-form). Such a property makes it an attractive target for crystallisation (Hughes, 2010), but no successful crystallisation of Cph1 in the Pfr-form has been reported to date.

A strategy is required to stabilise Agp1 in the Pfr-form in order to crystallise it. To this end a locked chromophore termed 15Ea was employed to obtain Agp1 in a stable Pfr-like form. 15Ea is a synthetic bilin with an additional carbon chain between the C and D rings to maintain the same configuration as that of BV in the Pfr-form (Section 1.3.3). The 15Ea adduct of Agp1 exhibits a similar absorption spectrum and kinase activity as the natural Pfr-form, and it is stable under illumination (Inomata et al., 2005), thus it was considered a good representative of the Pfr-form that is amenable to crystallisation.

A question is posed regarding the representativeness of the 15Ea adduct of Agp1 as the Pfr-form regardless of the length of the protein fragment. Recent evidences suggested the movement of the A-ring of the bilin moiety during the photoconversion process. A temperature-scan cryo-crystallographic study on PaBphP indicated the movement of the A-ring to follow the now undisputed C-D ring photoisomerisation (Yang et al., 2011). The A-ring of *SyB*-Cph1, a PAS-less phytochrome, was reported to be the major site of isomerisation during the photoconversion (Ulijasz et al., 2010). The BV in the Pr-forms features *syn* conformation at the C4-C5 single bond (Wagner et al., 2007, and Agp1 SER26-BV in the Pr-form). In 15Ea the configuration between the A and the B rings of 15Ea is unrestricted unlike doubly locked chromophores, therefore it is possible for the C4-C5 single bond between the A and the B-ring to change from the *syn* to the *anti* conformation if it is to occur during the formation of the Pfr-like form. A study with doubly locked chromophores could not unambiguously establish whether the conformation of the C4-C5 single bond in Agp1 is in *anti* or *syn* conformation (Inomata et al., 2009).

In this section, attempts at crystallographic structural determination of Agp1 in the Pfr-like form are described. A stable Pfr-like form was achieved by using the locked chromophore 15Ea, firstly with M20s followed by WT M15s and finally with SER mutants.

4.2 Analysis of the assembly process

4.2 1 Introduction

Incorporation of BV by the Agp1 apoprotein is characterised by efficient and rapid autolyase activity (Section 1.3.2 and 4.2, Lamparter et al., 2002). On the contrary, assembly with 15Ea takes longer, requiring approximately 140 minutes for completion (Inomata et al., 2005). The gradual rise of the extinction coefficient of the maximum in the Q-band was attributed to the formation of the covalent bond between the protein moiety and the 15Ea. In the studies of the assembly process presented in this section six combinations of holoproteins resulting from three different apoproteins and two chromophores were used, in order to empirically determine the time required for assembly to be complete. This is important, because sample homogeneity is a critical factor that influences the outcome of crystallisation experiments (Benvenuti and

Mangani, 2007). It is necessary to ensure that a high proportion of phytochrome is covalently bound to the bilin in order to achieve high sample purity and yields.

In the present work, the assembly process of a novel combination, between M20 and 15Ea, was also studied. Assembly product of M20 and 15Ea was hypothesised to resemble an intermediate state of photoconversion from the Pr- to the Pfr-form (T. Lamparter, personal communication). This was because the M20-BV adduct does not undergo full photoconversion due to the lack of the PHY domain (Noack et al., 2007). The stereochemistry of 15Ea around the C-D methine bridge mimics that of BV in the initial stage of photoconversion (Inomata et al., 2005). Assembly of M20 and 15Ea could in theory give rise to a holoprotein that represents the state immediately after the photoisomerisation of the C15=C16 double bond between the C and D rings, but without the subsequent thermal relaxation required for the formation of Pfr-form. It is difficult to predict the intermediate state which M20-15Ea would mimic, because the stage at which the photoconversion of M20-BV stalls is unclear. The problem also lies in currently ambiguous definitions of each intermediate state (Wagner et al., 2008).

Assembly processes of six types of holoproteins originating from combinations of three proteins (M15s, M15 C20A mutant, and M20s) and two bilin chromophores (BV and 15Ea) were initiated by mixing the protein and the bilin at a molar ratio of 1:1. Absorption spectra were taken immediately after the assemblies were initiated, followed by measurements at 15 minutes, 30 minutes, 45 minutes, 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 20.5 hours, and 21.5 hours. The absorption spectra are shown in Fig. 4.1. Assembly mixtures were incubated in darkness at 4 °C.

4.2.2 Holoprotein assemblies

Assembly processes of M15s-BV and M15 C20A-BV were rapid so that in each case no differences were observed in the absorption spectra immediately after and 15 minutes after the assembly initiation. Qualitatively similar rise of the Q-band peak were observed in both cases with the SAR value and λ_{max} being 1.20, 703 nm for M15s-BV and 0.79, 701 nm for M15 C20A-BV, respectively. Though a comparatively low extinction coefficient associated with the lack of covalent bond is in agreement with the previous observations (Lamparter et al., 2003, Lamparter et al., 2002), the blue-shift by 2 nm in λ_{max} of the non-covalent adduct was previously not recognised (Inomata et al.,

2005). This higher λ_{max} and absorption of the covalent adduct is likely to be the result of covalent bond formation, as in the case of assembly with 15Ea which is described below.

The slow assembly reaction of M15s-15Ea is in agreement with the previous observation (Inomata et al., 2005). The slow reaction was associated with a rise in the extinction coefficient of the peak in the Q-band and a red-shift of λ_{max}, and the reaction lasted up to 20.5 hours. No further change in the spectrum was observed thereafter. A slower reaction as found in the present study may be explained by the lower temperature (4 °C) of the assembly condition compared to 18 °C in the previous study (Inomata 2005).The absorption spectrum of the holoprotein exhibited λ_{max} in the Q-band at 740 nm with an SAR value of 0.58.

The assembly of M15 C20A-15Ea was already completed by the first measurement immediately after the initiation. The extinction coefficient at the Q-band maximum was lower (SAR: 0.22) and λ_{max} was blue-shifted (726 nm) compared to the covalently bonded adduct. This spectrum of the non-covalently bonded adduct is similar to the spectrum of M15s-15Ea obtained immediately after the reaction is initiated, prior to the absorbance increase and red-shifting of the peak in the Q-band (Inomata 2005). By using a mutant that lacks the necessary cysteine for covalent bond formation, I have confirmed that the slow reaction associated with increasing absorbance in the Q-band and red-shifting of λ_{max} is the result of covalent bond formation.

The assembly of M20s-BV was completed in 3 hours, and the spectrum is as described in Section 3.2.1.

The assembly of M20s-15Ea resulted in an unusual spectrum. The reaction was slow that it required up to 20.5 hours for completion. Two peaks with absorption maxima in the red (644 nm) and the far-red (714 nm) region appear immediately after the assembly is initiated, and the peaks in the red and far-red region blue-shift and red-shift respectively, such that the final λ_{max} are 638 nm and 727 nm. Extinction coefficients for each peak increase during the assembly, and the relative intensities of the peaks are reversed. The SAR of the peak in the red region at the start is 0.22 and finally reaches 0.38, but SAR values of the peak in the far-red region at the start and the end of the reaction are 0.17 and 0.52, respectively.

Despite the initial anticipation that M20s-15Ea may resemble a photoconversion intermediate state, the absorption spectrum of M20s-15Ea is very different from the meta-Rc (Fig. 4.2). The absorption spectrum of cryo-trapped Agp1 in the meta-Rc state is characterised by a peak approximately at 700 nm and a shoulder approximately at 750 nm (Borucki et al., 2005). Photoconversion of M20-BV results in a decrease of the absorbance of the peak in the Q-band, and the resulting spectrum was described as being similar to the spectrum of Agp1 in meta-Rc (Noack et al., 2007). The λ_{max} of the spectrum of M20s-15Ea and meta-Rc is very different, however. The initial λ_{max} of the peaks of M20s-15Ea are 644 nm and 714 nm. While differences in λ_{max} between holoproteins which have bound BV or a synthetic chromophore with corresponding stereochemistry have been noted previously (Inomata et al., 2005), the differences in this case are of greater magnitude (Table 4.1). A gradual increase in the λ_{max} and the absorption coefficient by the peak in far-red region of M20s-15Ea spectrum is qualitatively similar to the development of the Pfr-form during the natural photoconversion.

With regards to the increasing absorption at the far-red peaks in the natural Pfr-form and M20-15Ea, it is difficult to draw an analogy given the current lack of understanding in the molecular mechanism of photoconversion and a highly artificial combination of truncated M20s and the synthetic 15Ea chromophore. Overall it is difficult to relate the observed assembly of M20s-15Ea to the natural photoconversion of Agp1, and its significance was not conclusively validated in the present study.

	Natural, BV	***Synthetic, locked chromophores***
Pr-form	Agp1 full-length BV 702 nm (Inomata et al., 2005)	Agp1 full-length 15Za 714 nm (+12 nm) (Inomata et al., 2005)
Pfr-form	Agp1 full-length BV 747 nm (Inomata et al., 2005)	Agp1 full-length 15Ea 739 nm (- 8 nm) (Inomata et al., 2005)
Intermediate form	Agp1 full-length BV Approximately 700 and 750 nm (Borucki et al., 2005)	M20s-15Ea 644 nm (-56 nm) and 714 nm (- 36 nm)

Table 4.1 Forms of Agp1 protein and Q-band λ_{max} for each form, and the equivalent for the corresponding adduct with a synthetic chromophore. Differences in λ_{max} between the holoprotein with the natural chromophore and the corresponding synthetic chromophore are shown in parentheses next to the λ_{max} values of holoproteins with synthetic chromophore. The difference is relative to the λ_{max} of natural holoproteins. In cases of 15Za and 15Ea, the difference in λ_{max} from that of the corresponding natural holoprotein is smaller than the differences in λ_{max} of meta-Rc and M20s-15Ea, a holoprotein hypothesised to resemble an intermediate state.

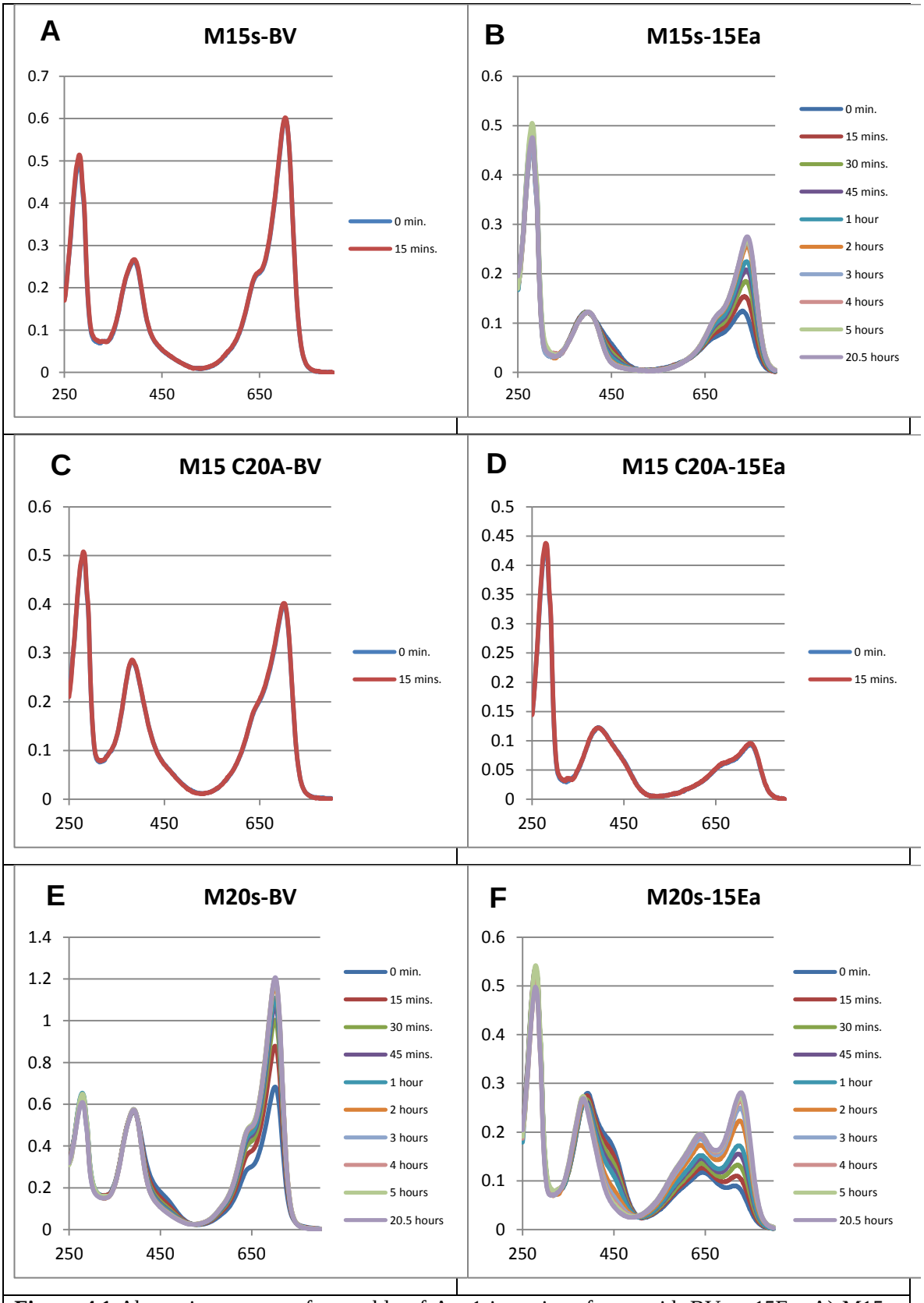


Figure 4.1 Absorption spectra of assembly of Agp1 in various forms with BV or 15Ea. A) M15s-BV. B) M15s-15Ea. C) M15 C20A-BV. D) M15 C20A-15Ea. E) M20s-BV. F) M20s-15Ea.

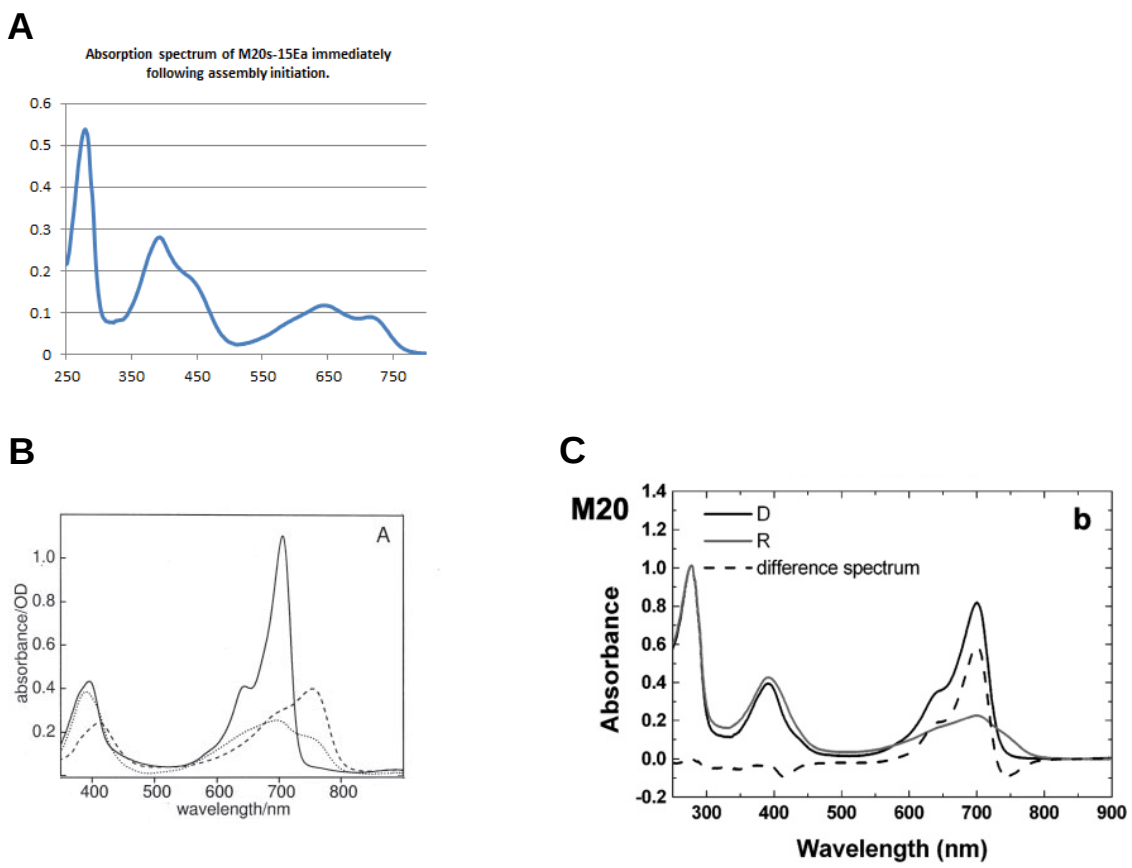


Figure 4.2 Absorption spectra of Agp1 holoproteins in an intermediate (-like) form. A) The spectrum of M20s-15Ea immediately after the assembly is initiated. B) The spectrum of Agp1 cryo-trapped in meta-Rc state (dotted line). Figure 6A from Borucki et al. (2005). C) The spectrum of photoconverted M20-BV (grey line). Figure 3 from Noack et al. (2007).

4.3 Large-scale production, analysis and crystallisation trials with the M20s-15Ea protein

Even though the relevance of any results obtained from M20s-15Ea for the naturally photoconverted holoprotein is uncertain, crystallisation trials were carried out in the hope to obtain some structural information. Structural information of M20, a truncated construct, may not result in the full elucidation of the mechanism of signal transduction, but may show the altered protein-chromophore interaction in the vicinity of the bilin. The plausibility of such information may be validated by comparing it with the structure of PaBphP in the Pfr-form.

4.3.1 Expression and purification of M20s-15Ea protein

Affinity purified M20s apoprotein (Section 3.3.2) was used to produce M20s-15Ea holoprotein. 20 mg of apoprotein was mixed with 15Ea in DMSO at a molar ratio of 1:1 under the presence of 10 mM DTT.

Following 18 hours of incubation in darkness, the assembly mixture was dialysed against the IEX Buffer A (Section 2.2.3.7.4). A linear gradient elution from IEX Buffer A to IEX Buffer B was used to elute the M20s-15Ea protein from a HiTrap Q HP column (Section 2.2.3.7.4). Three peaks were eluted, of which one of them appears as a shoulder of the main peak. Fractions containing each peak were analysed by absorption spectroscopy (Fig. 4.3). The spectra of two minor peaks sampled at fractions C6 and C10 resemble spectra observed at later and early stages of the assembly of M20s-15Ea, respectively (Section 4.2.2). Interestingly, the spectrum of the main peak sampled from fraction C1 appears distinctly different from that of C6 and C10, because a peak in the far-red region with a λ_{max} of 727 nm with a shoulder approximately at 650 nm is found, which is similar to the Pfr-form. Amount of protein in each peak occur in decreasing order of, main peak (C1, Pfr-like), first minor peak (C6, later stage of assembly), and second minor peak (C10, early stage of assembly), as judged from the chromatogram.

IEX chromatography has revealed the presence of at least three products for the assembly of M20s protein with 15Ea which could be separated by IEX. Emergence of a species with a Pfr-like spectrum in the Q-band was undetected in the earlier test (Section 4.2.2) most likely due to the persisting presence other two species with different absorption spectra. In order to confirm that the species from two minor peaks

were not the intermediates of a slow assembly process, fractions C6 and C10 were incubated in darkness for 48 hours while the absorption spectra were measured at 24 hour intervals (Fig. 4.4). Absorption spectra of fraction C6 and C10 slightly changed following incubation for 24 hours. The change is characterised by an increase in the extinction coefficient of the far-red peak, and this is in agreement with the previous observation (Section 4.2.2). No further change was observed in the spectra in the following 24 hours, and both spectra retained two absorption peaks in the Q-band, thus confirming that the species from two minor peaks are not assembly intermediates, but alternative assembly products.

It is unclear why certain fractions of the protein results in the formation of products with different spectral properties. This may be explained by the inhomogeneity in the protein sample or 15Ea chromophore which leads to several species of holoproteins.

Fractions containing the main peak eluted during an IEX were pooled and purified using an isocratic elution form a Superdex 200 column using SEC buffer with no NaCl (Section 2.2.3.7.6). Two peaks at elution volumes of 14.95 and 19.77 ml were obtained, and absorption spectra of two eluted peaks appear almost identical (Fig. 4.5). The broadness of the main eluted peak was attributed to the high concentration of the injected sample. The higher elution volume of the minor peak with a higher elution volume was attributed to the result of non-specific interaction rather than the difference in oligomeric state as judged from the calibration of the S200 column (Appendix 5). As a precaution only the fractions containing the main peak (fractions A12-B11) were pooled and concentrated for crystallisation trials, however.

The absorption spectrum of the purified sample has a λ_{max} in the Q-band at 727 nm and an SAR value of 0.69. Similarly to the case with M20-BV, a low SAR value is attributed to the lack of the PHY domain (Section 3.2.1).

Assembly of M20s-15Ea ultimately yielded holoprotein with an absorption spectrum similar to that of the Pfr-form. Various forms of M20s protein were analysed by a native-PAGE gel which confirmed the conformational differences between the forms of M20s protein, in agreement with the IEX (Fig. 4.6).

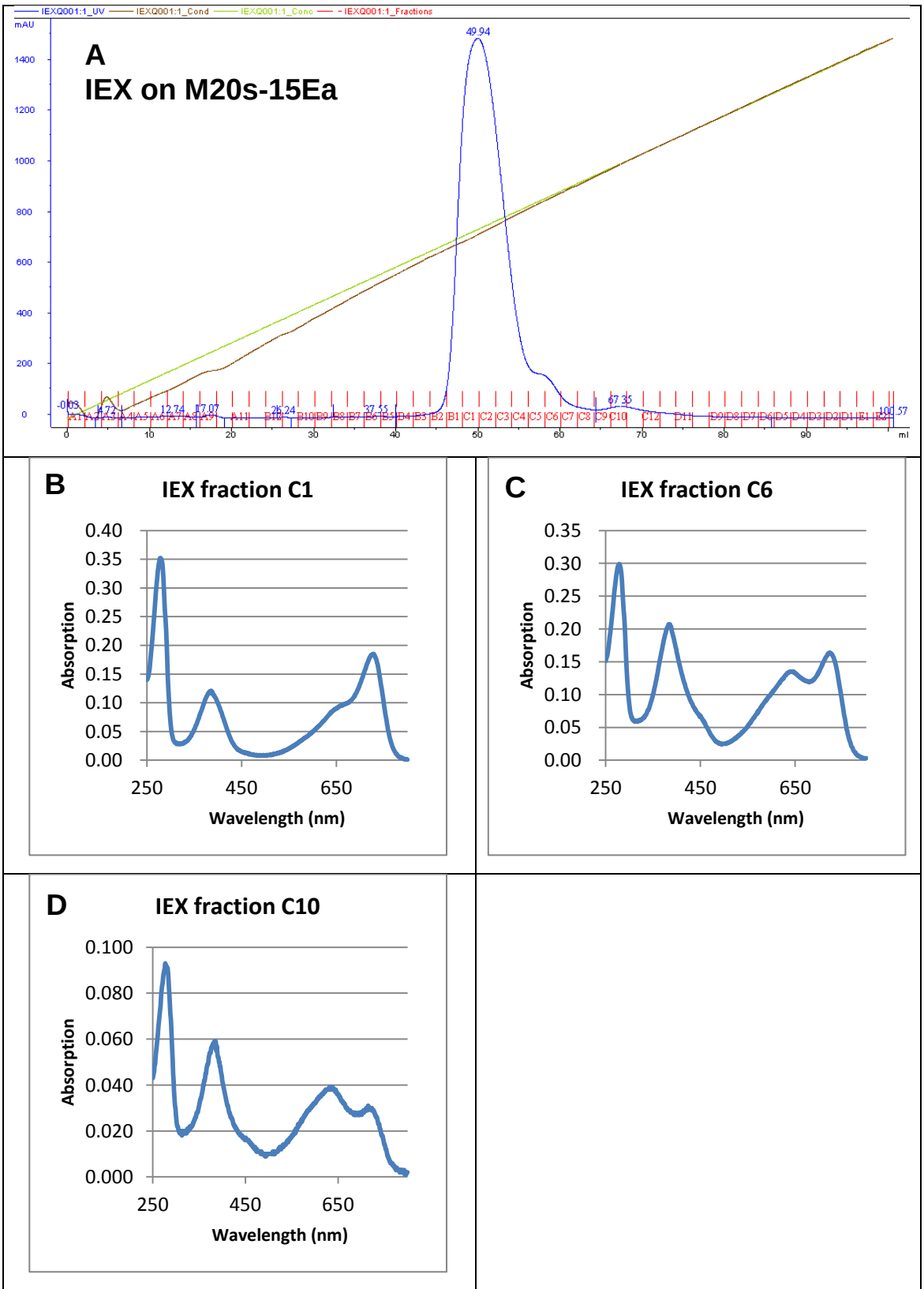


Figure 4.3 Purification of M20s-15Ea assembly product with IEX. A) IEX elution profile of M20s-15Ea during linear gradient elution, resulting in three peaks with distinct absorption spectra. The blue, brown, and green lines indicate the absorption (mAU), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively. UV-vis absorption spectra of B) Fraction C1, C) Fraction C6, and D) Fraction C10. The sample from fraction C1 was diluted by 1/8 with IEX buffer A prior to the measurement.

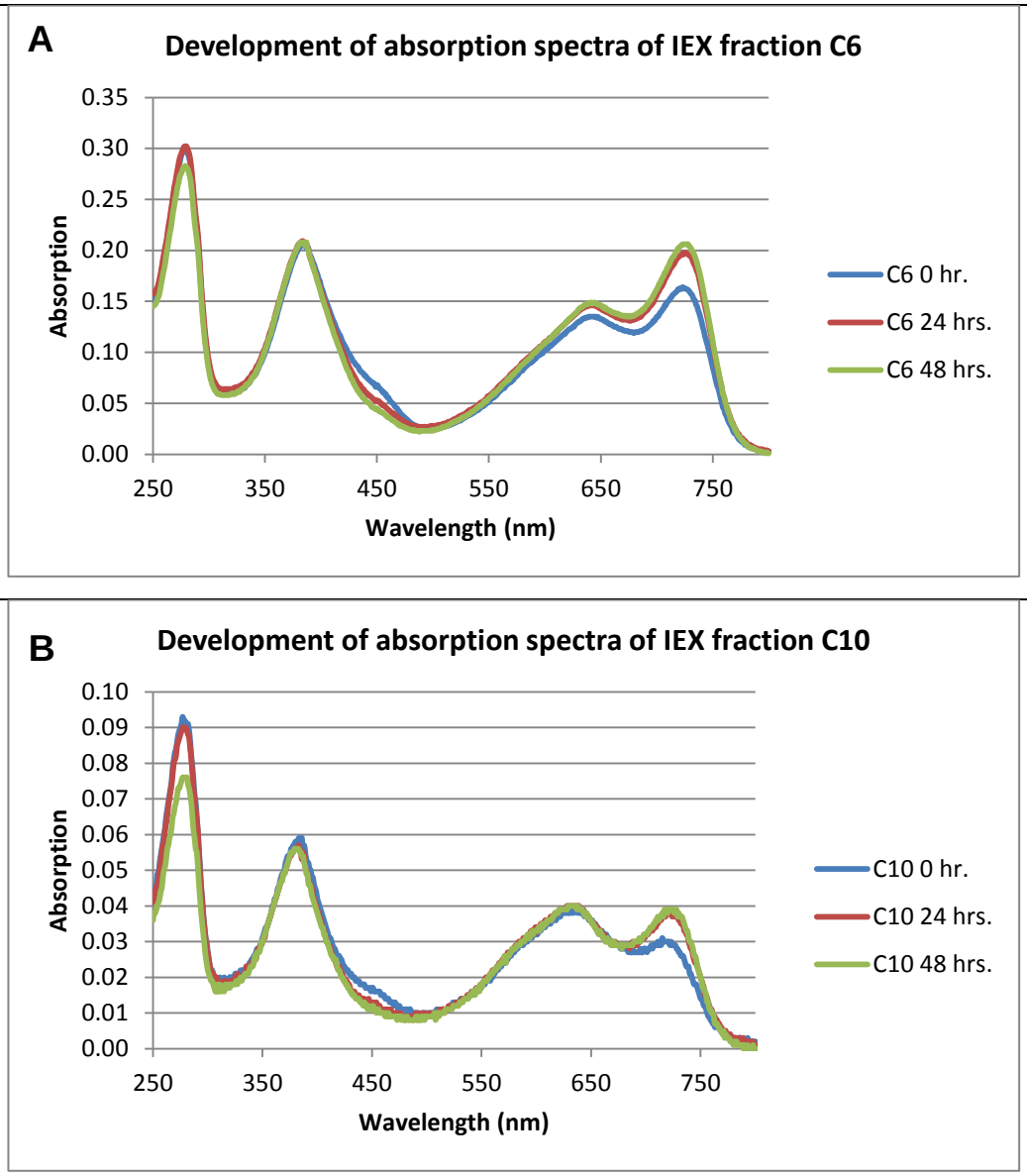


Figure 4.4 UV-vis absorption spectra of species from two minor peaks separated by IEX. A) Absorption spectra of fraction C6 and B) C10 immediately after separation by IEX and after incubation for 24 hours, followed by additional 24 hours of incubation.

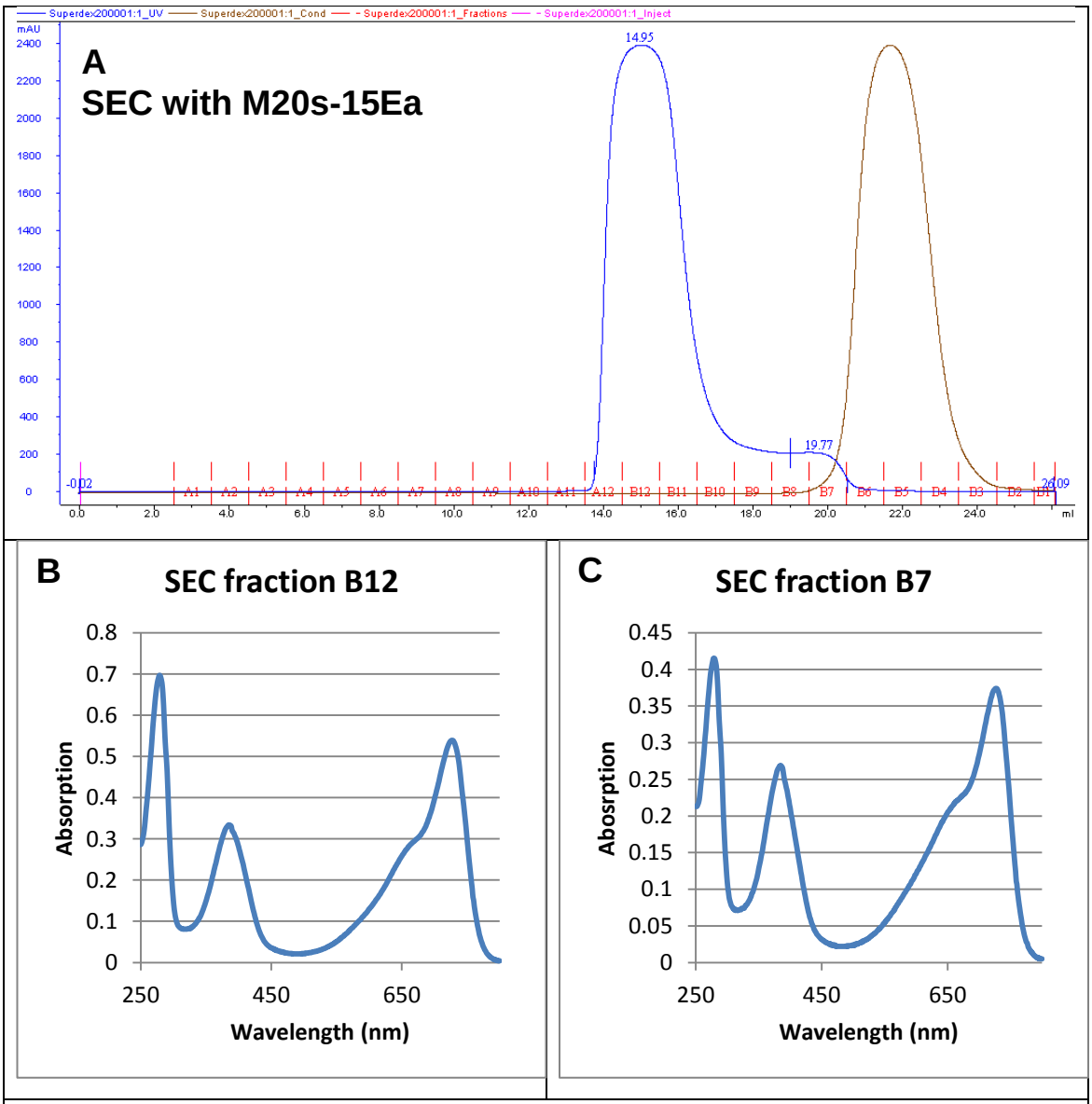


Figure 4.5 Purification of M20s-15Ea with SEC. A) SEC elution profile of M20s-15Ea obtained by isocratic elution. The blue and brown lines indicate the absorption (mAU at 280 nm) and conductivity (mS/cm), respectively. UV-vis absorption spectra of fractions, B) Fraction B12; C) Fraction B7.

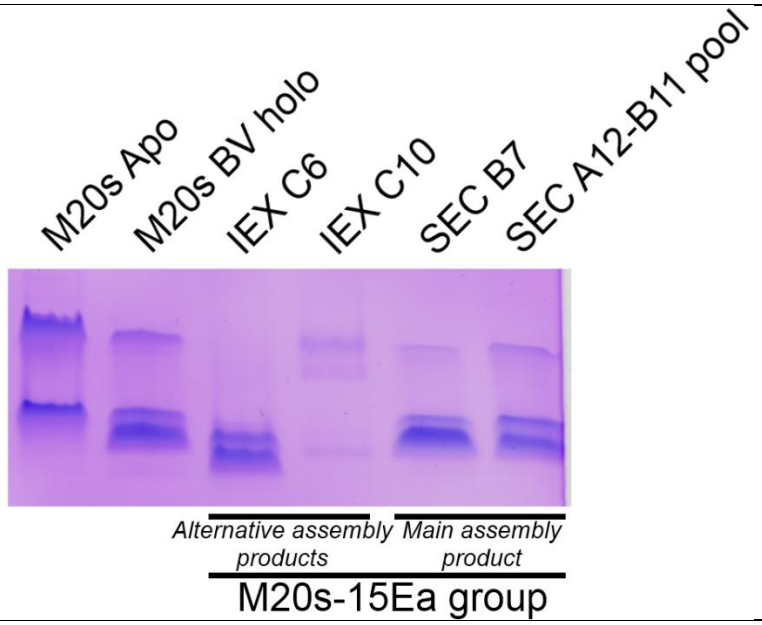


Figure 4.6 A native-PAGE gel showing migration patterns of different forms of M20s protein. The last two lanes corresponding to the two peaks from SEC which were deemed identical as judged by absorption spectra (Fig. 4.4). Identical absorption spectra agree with the identical migration patterns shown here. The migration patterns of the main assembly product of M20s-15Ea in the last two lanes are similar to that of M20-BV, but distinctly different from that of M20 apoprotein. The migration patterns of the two alternative assembly products were different from each other, as well as any other forms of M20s.

4.3.2 Crystallisation screening with purified M20s-15Ea protein in a Pfr-like state

Purified Pfr-like M20s-15Ea holoprotein was subjected to crystallisation trials with JCSG+, PACT-premier, PGA, and Morpheus kits. 2-drop 96-well MRC plates with 80 µl reservoirs were used and drops comprised 200 nl of protein solution at 25 mg/ml and 200 nl of reservoir solution. The plates were incubated at 19 °C in darkness, and were subjected to regular inspection, but no lead condition was identified.

4.4 M15s-15Ea: Pfr-like form

No lead crystallisation condition was identified for the M20s-15Ea holoprotein, but no further efforts were made for the following reasons. Firstly, the M20 protein lacks the PHY domain, therefore its structure is not ideal for gaining functional insights of the protein. Secondly, the absorption spectrum of M20s-15Ea is more different from that of M15s-15Ea compared to the native absorption spectrum of Agp1 in the Pfr-form (Fig. 4.5). In turn more efforts were made to solve the structure of the 15Ea adduct of the M15s protein to solve the structure of Agp1 in the Pfr-like form that represents the native Pfr-form more closely and provides deeper functional insights.

The assembly process of M15s-15Ea takes a longer time than it does in the case of M15s-BV, as the process involves the quick initial binding followed by slower formation of the covalent bond (Section 4.2.2, Inomata et al., 2005). For this reason, assembly of M15s-15Ea was carried out by a long incubation in darkness similarly to the cases of M20(s)-BV and M20s-15Ea, because a longer assembly process increases the conformational homogeneity of the sample.

Additionally M15s-15Ea was further purified by IEX to separate the holoprotein from the apoprotein. The differences in elution volumes of the apoprotein and the holoprotein is less pronounced in M15 proteins than in M20 proteins (Section 6.2), but these differences were sufficient for separation of the proteins and thereby to improve the purity of M15-15Ea. HIC was not used since no difference in elution volumes of apoprotein and holoproteins of M15 proteins was observed (Section 6.4.1).

4.4.1 Expression and purification of M15s-15Ea protein

Affinity purified M15s apoprotein (Section 3.4.1) was assembled with 15Ea by adding 15Ea stock solution in DMSO to a molar protein/chromophore ratio of 1. The assembly was allowed to continue in darkness for 18 hours at 4 °C under the presence of 10 mM DTT. The sample was then dialysed against IEX buffer A. A linear gradient elution from IEX Buffer A to IEX Buffer B was used to elute the M15s-15Ea protein from a HiTrap Q HP column (Section 2.2.3.7.4), resulting in elution of three peaks (Fig. 4.7). Two minor peaks with larger elution volumes were judged to consist of holoprotein with different conformation from that of the protein in the main peak with a Pfr-like absorption spectrum. Fractions containing holoprotein with similar spectral properties to the previously published ones (D8 – D1) were pooled, concentrated, and purified by SEC (Fig. 4.8). Isocratic elution yielded a main elution peak and a minor peak with a larger elution volume. The latter was attributed to non-specific interaction between the protein and the matrix, rather than protein in different oligomeric state as judged from the calibration of the S200 SEC column (Appendix 5). As a precaution, only the fractions of the main peak were pooled, and the holoprotein was concentrated to typically 30 mg/ml. The SAR value of the purified M15s-15Ea holoprotein was usually about 0.90, which is higher than any published results, and it is an indication of more successful purification of holoprotein by utilising IEX.

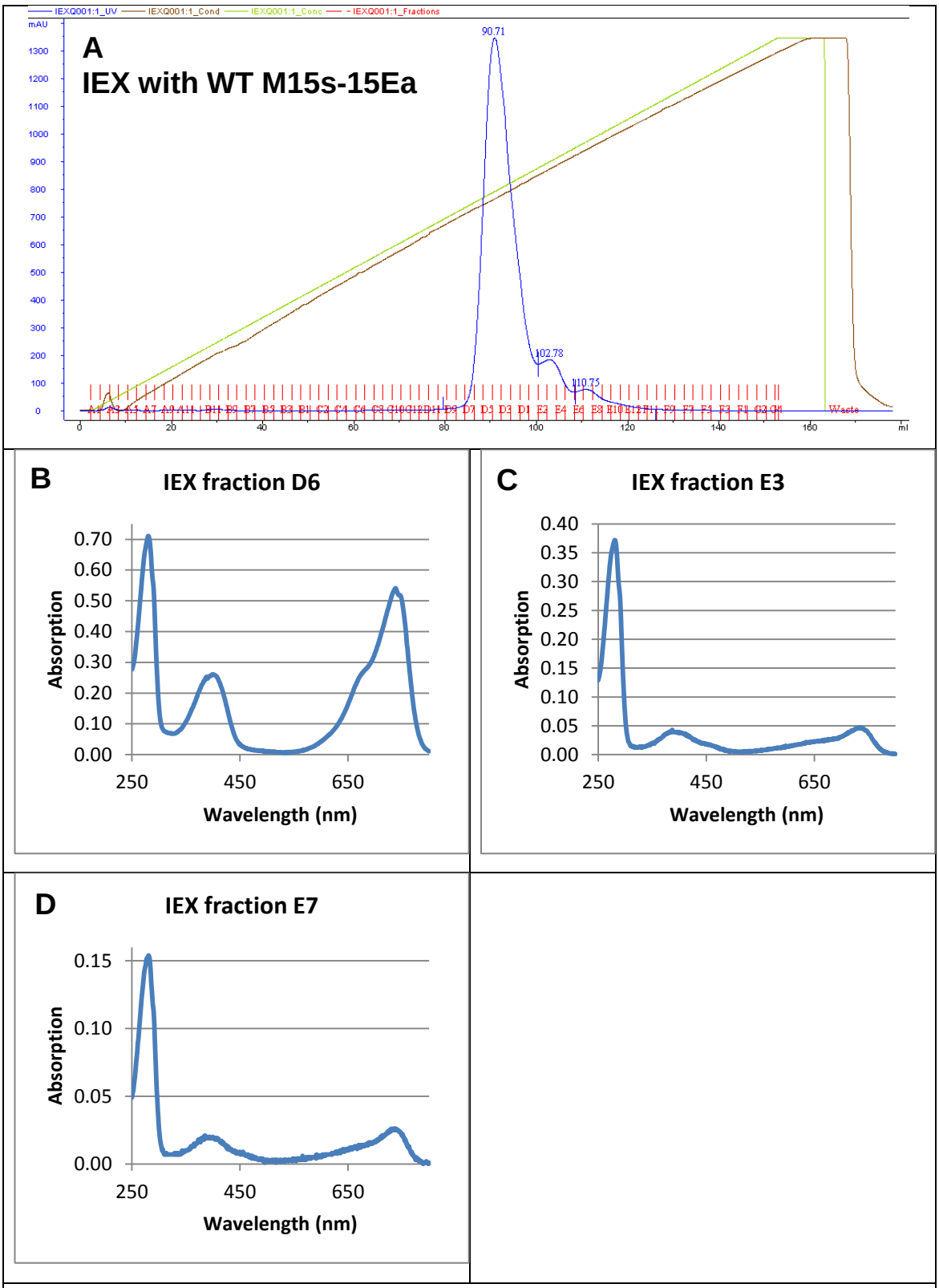


Figure 4.7 Purification of the M15s-15Ea holoprotein with IEX. A) IEX elution profile of M15s-15Ea from a linear gradient elution. The blue, brown, and green lines indicate the absorption (mAU at 280 nm), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively. UV-vis absorption spectra of fractions, B) Fraction D6, SAR = 0.90; C) Fraction E3, SAR = 0.13; D) Fraction E7, SAR = 0.17. Even though the spectra of the two minor peaks indicate the presence of the chromophore in the protein, the low absorbance in the Q-band which is unlike the natural Pfr-form might reflect conformational heterogeneity of the protein sample.

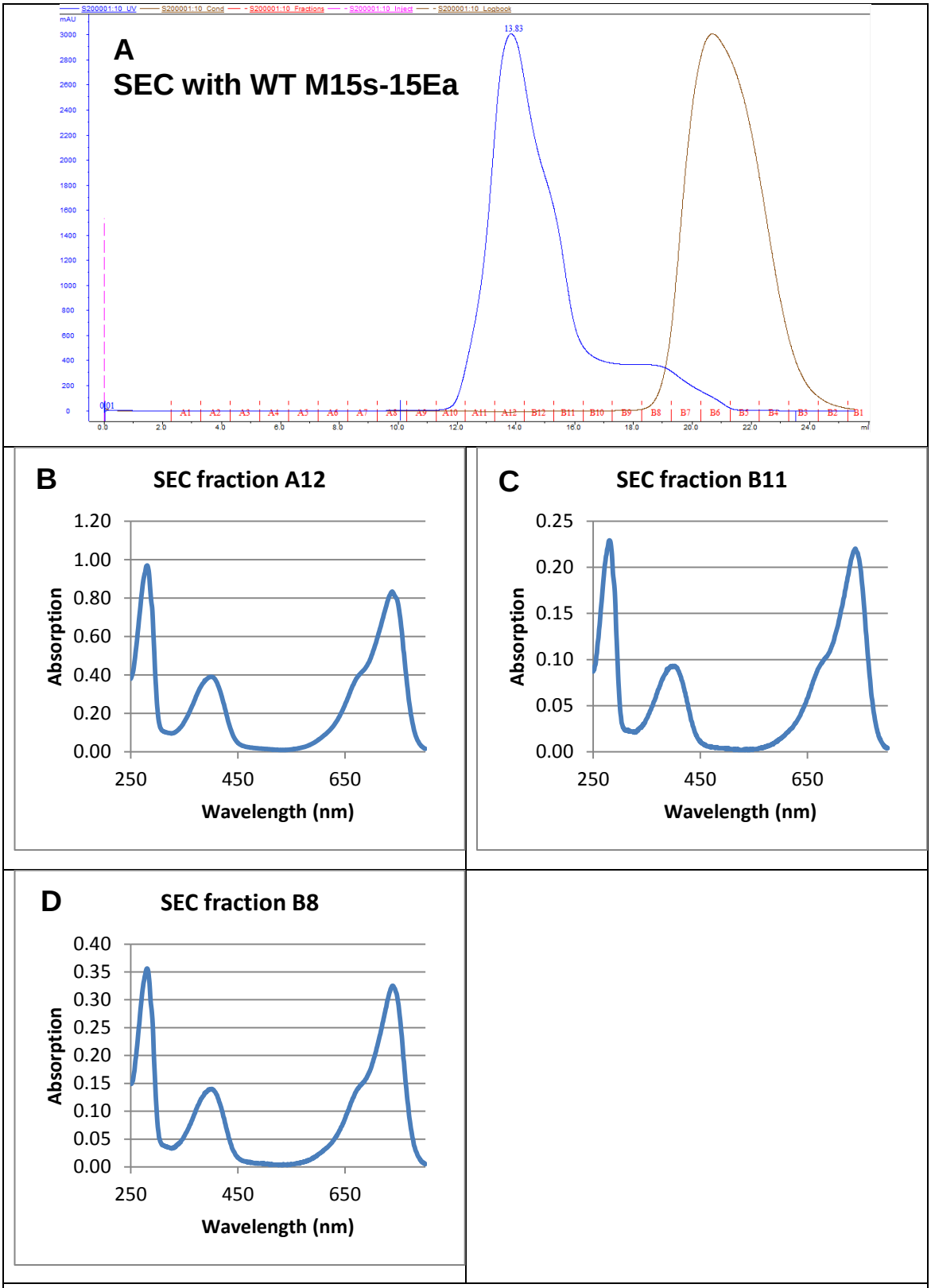


Figure 4.8 SEC of M15s-15Ea holoprotein. A) Chromatogram of isocratic elution. The blue and brown lines indicate the absorption (mAU at 280 nm) and conductivity (mS/cm), respectively. UV-vis absorption spectra of fractions from SEC. B) Fraction A12, SAR = 0.86; C) Fraction B11, SAR = 0.96; D) Fraction B8, SAR = 0.91. All the protein-containing fractions exhibited essentially identical absorption spectra. Larger elution volume by the second peak was judged to result from non-specific interaction between protein and the matrix.

4.4.2 Crystallisation screening with purified M15s-15Ea protein

Purified M15s-15Ea holoprotein was subjected to crystallisation trials with JCSG+, PACT-premier, PGA, and Morpheus kits (Molecular Dimensions). 2-drop 96-well MRC plates with 80 µl reservoirs were used and drops comprised 200 nl of protein solution at 25 mg/ml and 200 nl of reservoir solution. The plates were incubated at 19 °C in darkness, and plates were routinely inspected (Section 2.2.6.1), but no lead condition was identified. Furthermore, crystallisation of M15s-15Ea was attempted under a condition which had been successfully used for M15s-BV holoprotein (16% PEG3350, 200 mM $CaCl_2$, 100 mM Tris-HCl pH 8.0 at 20 °C), but no crystallisation was observed.

4.5 SER mutants-15Ea

Screening the 15Ea adducts of WT M15s and M20s proteins did not result in identification of a lead crystallisation condition. The failure to obtain M15s-15Ea crystal under the crystallisation condition identified for M15s-BV suggests the conformation of the Pr-form is different from that of the Pfr-like form, so that identical crystal contacts cannot occur to form crystals of the same form.

The structure of Agp1 in the Pr-form was determined as a result of SER mutation strategy (Chapter 3). The success was attributed to the introduction of mutations which resulted in the crystal form to change from the “pathological” form to the one with better diffraction quality. Since SER mutation patches seem to have contributed to the formation of novel crystal forms in the case of SER26-BV (Section 5.5), it was hypothesised that the SER strategy may also increase the probability of the 15Ea adduct to crystallise by providing a novel contact surface for crystallisation to occur.

To this end, all the SER mutants described so far, SER2, SER3, SER6, SER26, and SER36, were assembled with 15Ea and subjected to crystallisation trials.

4.5.1 Expression and purification of 15Ea adducts of SER mutants

Apoproteins of SER mutants (SER2, SER3, SER6, SER26, and SER36) were expressed and affinity purified as described previously (Section 3.6.3 and 3.7.3). Affinity purified SER apoproteins at typical concentration of 10 mg/ml were assembled with 15Ea following the identical protocol as for the WT M15 (Section 4.4.1). No difference in assembly characteristics were observed for the different mutants adducts. 15Ea adducts

of the SER mutants were tested by several biochemical methods to prove that functions of the mutants are not inadvertently affected (Section 5.3).

Since the assembly characteristics of SER mutants with 15Ea did not differ from that of the WT, IEX purification was necessary to purify the properly folded holoprotein with the Pfr-like absorption spectrum. All the 15Ea adducts of the SER mutants were subjected to IEX and SEC as described in Section 4.4.1, resulting in holoproteins with spectra that are indistinguishable from that of the WT M15s-15Ea (SER2-15Ea, Fig. 4.9; SER3-15Ea, Fig. 4.10, SER6-15Ea, Fig. 4.11; SER26-15Ea, Fig. 4.12; SER36-15Ea, Fig. 4.13).

In all cases elution profiles were essentially identical to one another throughout the purification protocol. This shows that physical properties, as indicated by similar elution volumes in IEX and SEC, of the SER mutants have not been majorly changed (Table 4.2).

Elution profiles from SEC are characterised by a single peak for SER2-15Ea and SER3-15Ea, whereas two peaks are detected for SER6-15Ea, SER26-15Ea, and SER36-15Ea. The second lower peak cannot be holoprotein in an alternative oligomeric state according to the calibration of the S200 column (Appendix 5). Instead non-specific interactions between the protein and the stationary phase were suspected to be the reason for the occurrence of these peaks. In all cases the absorption spectra of the two peaks have almost identical SAR values, but as a precaution only the fractions containing the main peak were pooled and concentrated for crystallisation trials.

This larger elution volume of the 15Ea adduct is exclusive to those holoprotein with the SER6 mutation cluster, therefore hydrophobic interaction was initially assumed to occur at the SER6 mutation site (E336A/K337A). Yet the elution profile of WT M15s-15Ea and M20s-15Ea holoproteins also exhibited similar delayed second peak (Fig. 4.7 and Fig. 4.8, respectively), and M20s does not even have a PHY domain where SER6 mutation site resides. The cause for this seemingly unpredictable appearance of delayed elution in SEC was not investigated further in this study.

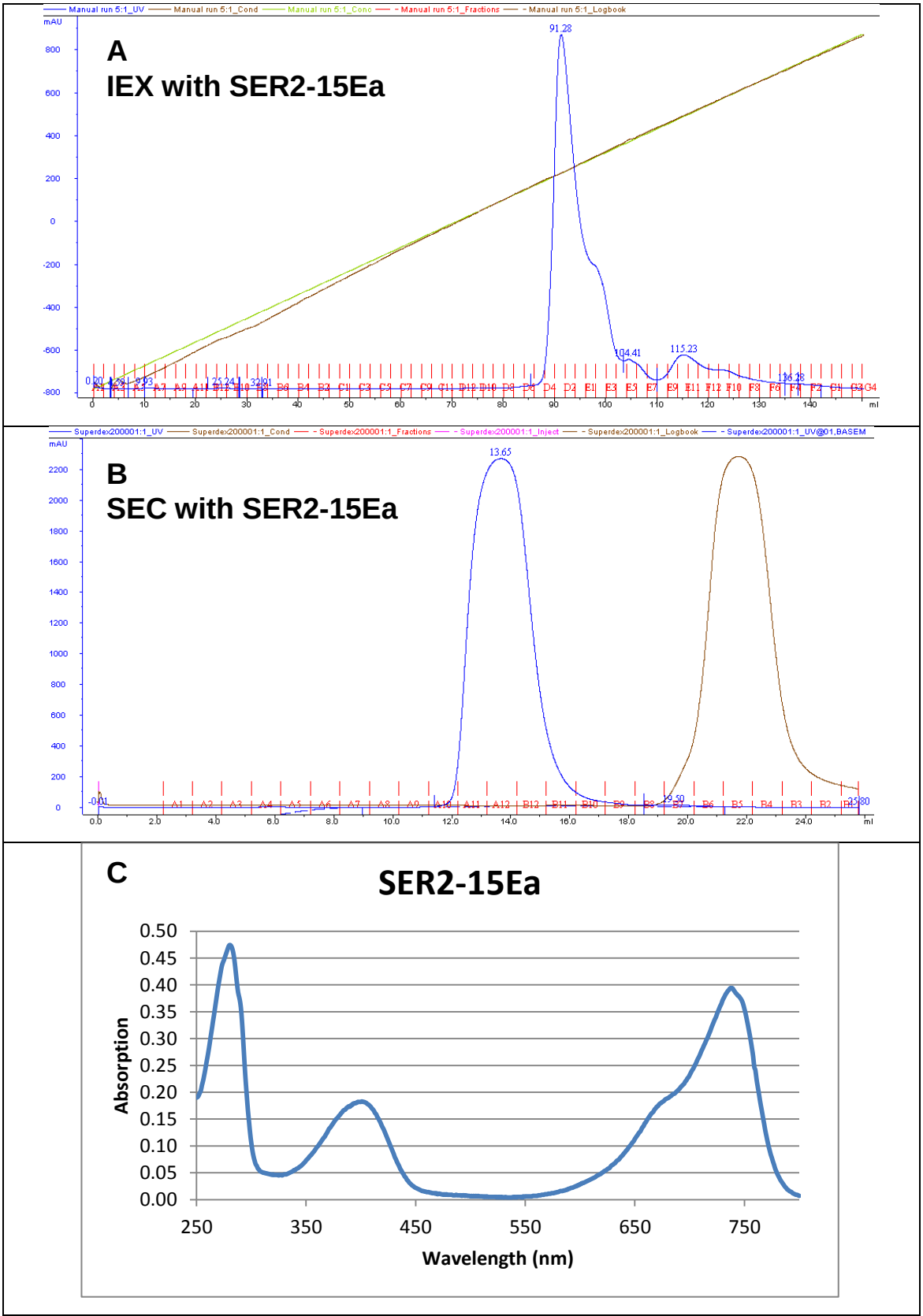


Figure 4.9 Purification with IEX and SEC and a UV-vis absorption spectrum of the purified SER2-15Ea holoprotein. A) The chromatogram from IEX. B) The chromatogram from SEC. The blue, brown, and green lines indicate the absorption (mAU at 280 nm), conductivity (mS/cm), and proportion buffer A/B (%), respectively. C) An absorption spectrum of the pooled SER2-15Ea holoprotein.

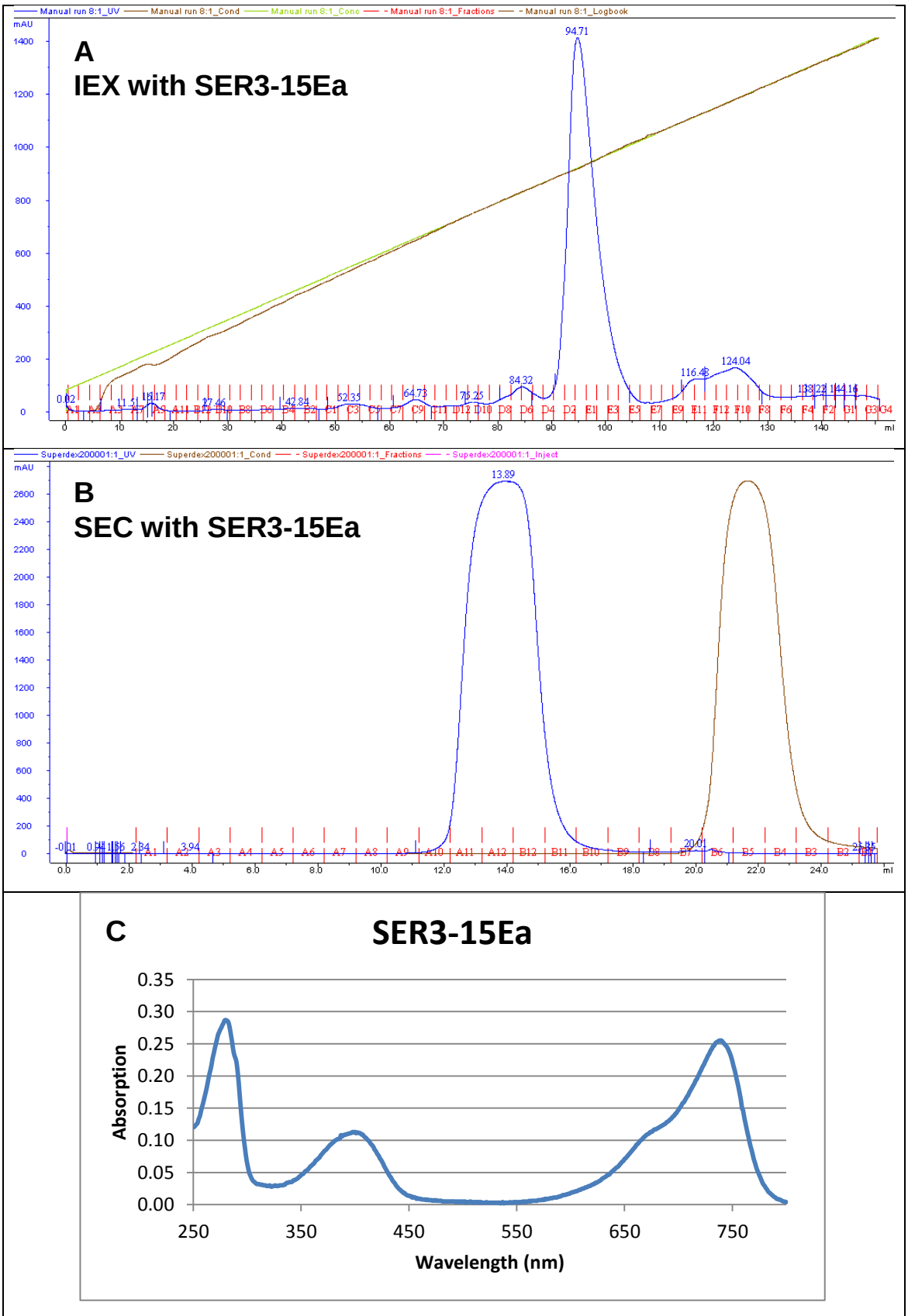


Figure 4.10 Purification with IEX and SEC and a UV-vis absorption spectrum of the purified SER3-15Ea holoprotein. A) The chromatogram from IEX. B) The chromatogram from SEC. The blue, brown, and green lines indicate the absorption (mAU at 280 nm), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively. C) An Absorption spectrum of the pooled SER3-15Ea holoprotein.

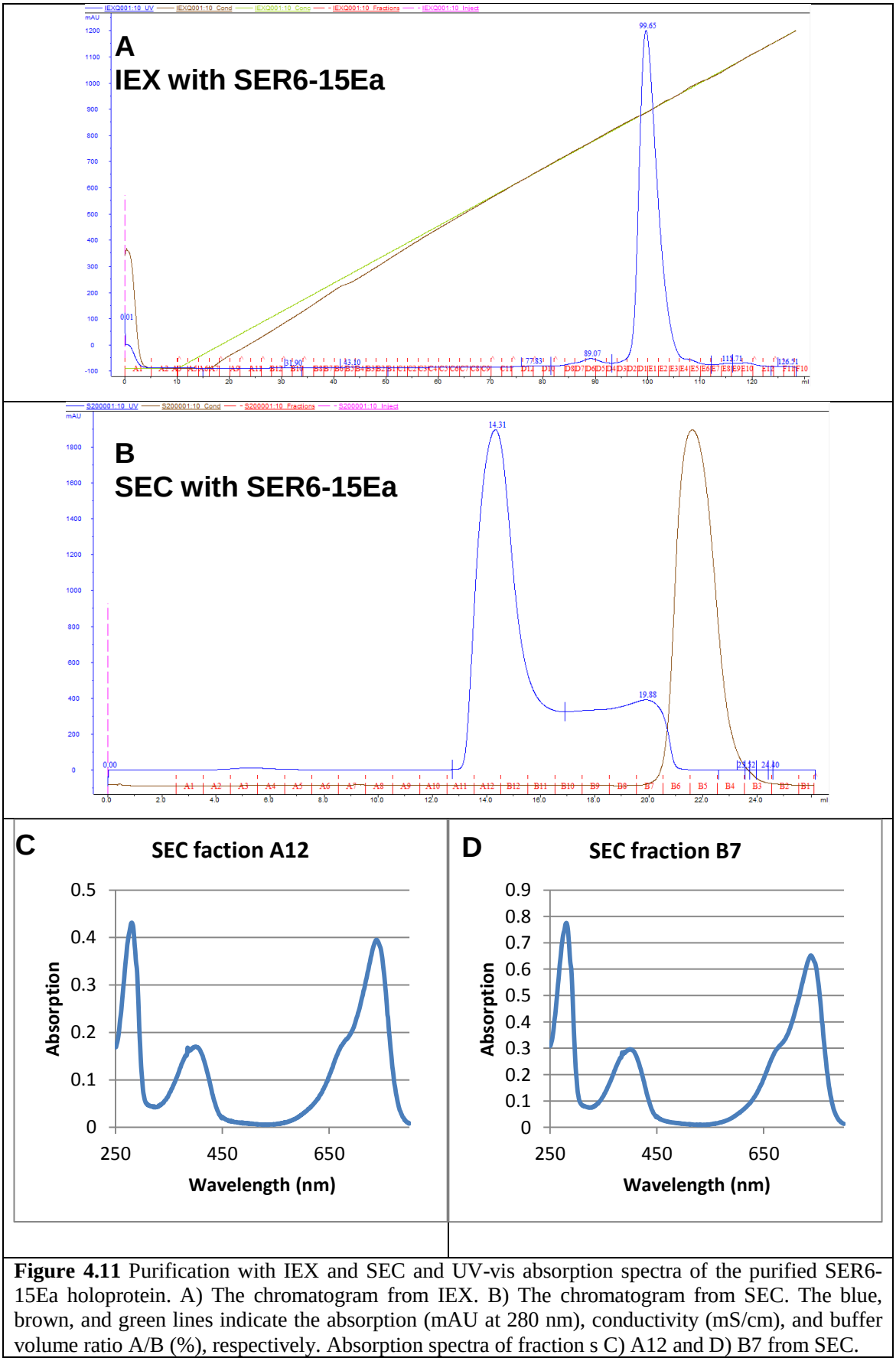


Figure 4.11 Purification with IEX and SEC and UV-vis absorption spectra of the purified SER6-15Ea holoprotein. A) The chromatogram from IEX. B) The chromatogram from SEC. The blue, brown, and green lines indicate the absorption (mAU at 280 nm), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively. Absorption spectra of fraction s C) A12 and D) B7 from SEC.

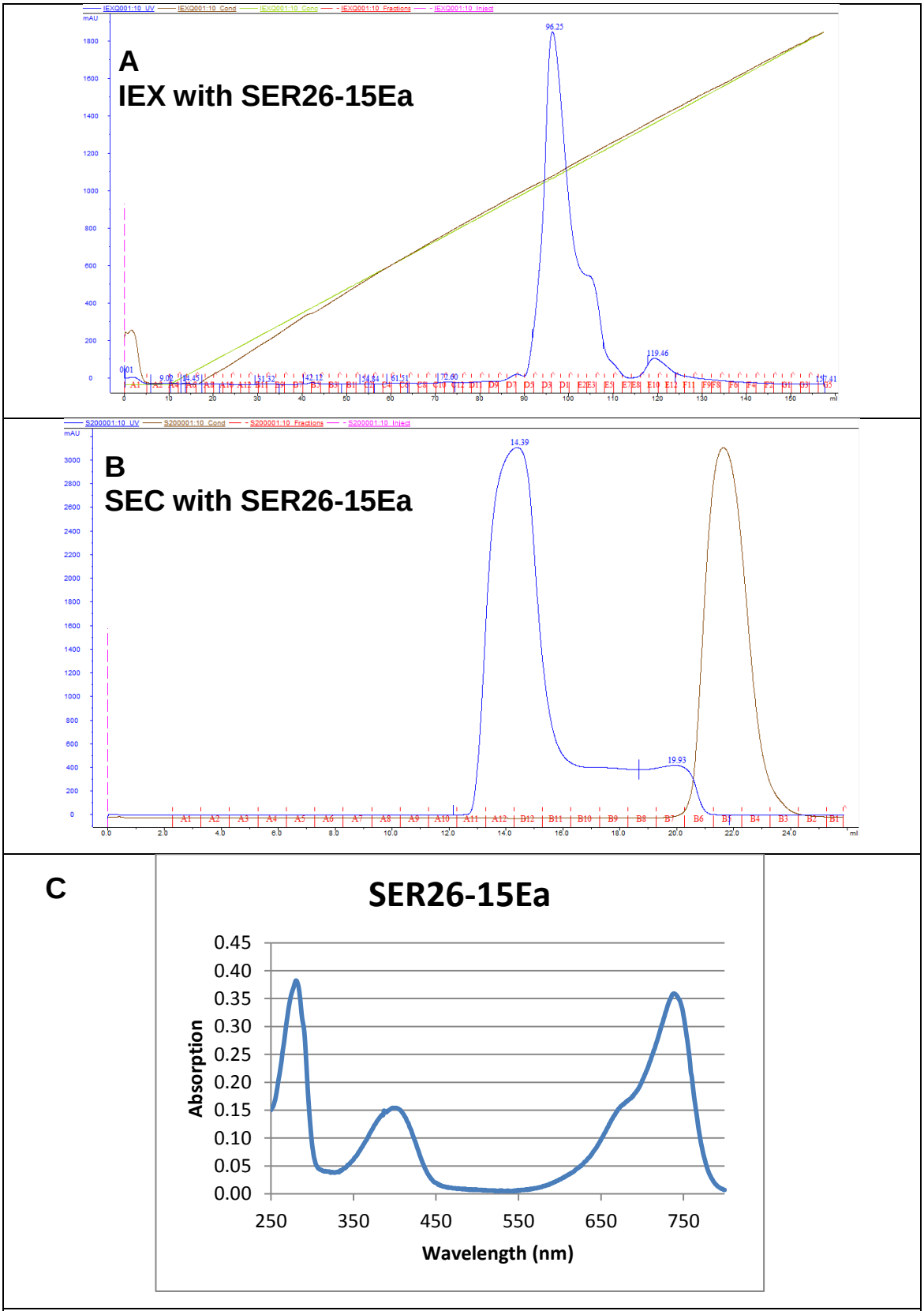


Figure 4.12 Purification with IEX and SEC and a UV-vis absorption spectrum of the purified SER26-15Ea holoprotein. A) The chromatogram from IEX. B) The chromatogram from SEC. The blue, brown, and green lines indicate the absorption (mAU at 280 nm), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively. C) An absorption spectrum of the pooled SER26-15Ea holoprotein.

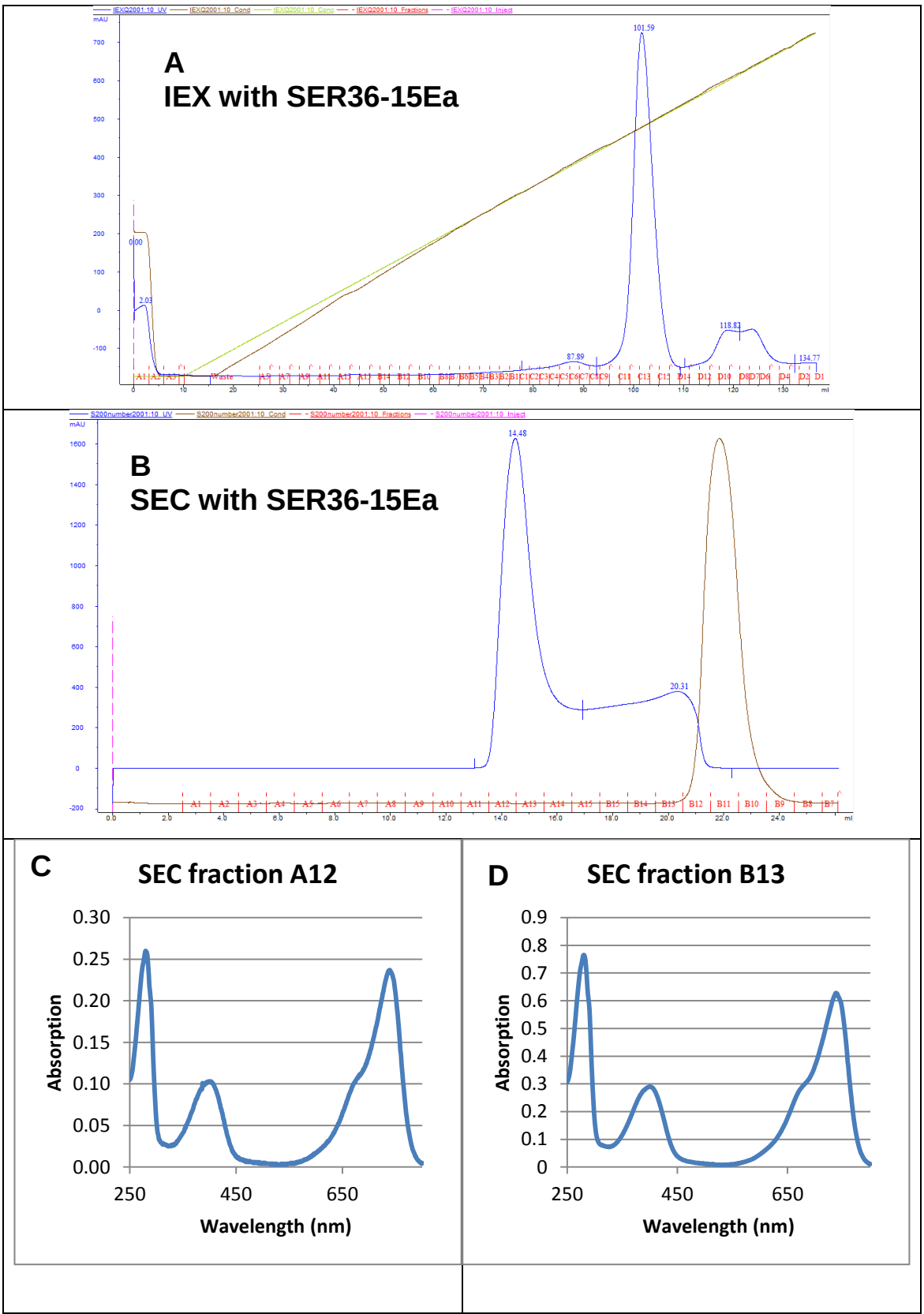


Figure 4.13 Purification with IEX and SEC and UV-vis absorption spectra of the purified SER36-15Ea holoprotein. A) The chromatogram from IEX. B) The chromatogram from SEC. The blue, brown, and green lines indicate the absorption (mAU at 280 nm), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively. Absorption spectra of fractions C) A12 and D) B13 of SEC on SER36-15Ea.

Protein	IEX elution volume (ml)	SEC elution volume (ml)
M15s-15Ea		13.83
SER2-15Ea	91.28	14.29
SER3-15Ea		13.89
SER6-15Ea	99.65	14.31
SER26-15Ea	96.25	14.39
SER36-15Ea	101.59	14.48

Table 4.2 Similar elution volumes of the 15Ea adducts of the SER mutants in both IEX and SEC. Elution volumes where purification was carried out under the identical protocol were collected for both IEX and SEC. Entire elution volumes for IEX and SEC were 150 ml and 24 ml, respectively. Standard purification protocol for IEX was a linear gradient from IEX Buffer A to IEX Buffer B over 30 CV at 0.5 ml/min, and an isocratic elution over 1 CV at 0.5 ml/min for SEC.

4.5.2 Crystallisation screening with purified SER mutants-15Ea protein

Purified 15Ea adducts of SER mutants were subjected to crystallisation trials with JCSG+, PACT-premier, PGA, and Morpheus kits (Molecular Dimensions). 2-drop 96-well MRC plates with 80 μl reservoirs were used and drops comprised 200 nl of protein solution at 25 mg/ml and 200 nl of reservoir solution. The plates were incubated at 19 °C in darkness, and were regularly inspected.

Screening resulted in identification of several lead conditions for SER2-15Ea using the Morpheus kit. These conditions are listed in Table 4.3. The optimised crystallisation condition consisted of 0.07 M D-galactose, 0.1 M MES/Imidazole pH 6.5, 20% Glycerol and 10% PEG4000. Hexagonal rod shaped crystals typically grew to dimensions of ~ 0.3 mm × 0.15 mm × 0.15 mm in 2 weeks of incubation in darkness. For preparation purposes, crystallisation was carried out in a 24-well plate by the hanging drop method with 400 μl of reservoir solution. All crystals grown in conditions from or derived from the Morpheus kit were vitrified in liquid nitrogen using LithoLoops without further cryoprotection, because the solutions contain 20% cryoprotectant.

Often rod-shaped crystals were found to possess re-entrant faces at the longitudinal termini (Fig. 4.14C). This is a sign of twinning and should be noted for potential difficulties in processing the dataset (Hammond, 2009: 21-25).

Condition	Additives	Buffer	Cryoprotectant	Precipitant
B3	0.03 M Sodium bromide 0.03 M Sodium fluoride 0.03 Sodium iodide	0.1 M MES/Imidazole pH 6.5	20% Glycerol	10% PEG4000
F3	0.02 M D-galactose 0.02 M D-glucose 0.02 M D-mannose 0.02 M D-xylose 0.02 M L-fucose 0.02 M N-acetyl-D-glucosamine	0.1 M MES/Imidazole pH 6.5	20% Glycerol	10% PEG4000
G2	0.02 M Ammonium acetate 0.02 M Sodium citrate 0.02 M Sodium formate 0.02 M sodium oxamate 0.02 M Sodium potassium tartate	0.1 M MES/Imidazole pH 6.5	20% Ethylene glycol	10% PEG8000
H6	0.02 M Alanine 0.02 M glutamic acid 0.02 M Glycine 0.02 M Lysine HCl 0.02 M Serine	0.1 M MOPS/Sodium Hepes pH 7.5	20% Ethylene glycol	10% PEG8000

Table 4.3 Lead conditions identified using the Morpheus kit in a crystallisation trial with SER2-15Ea.

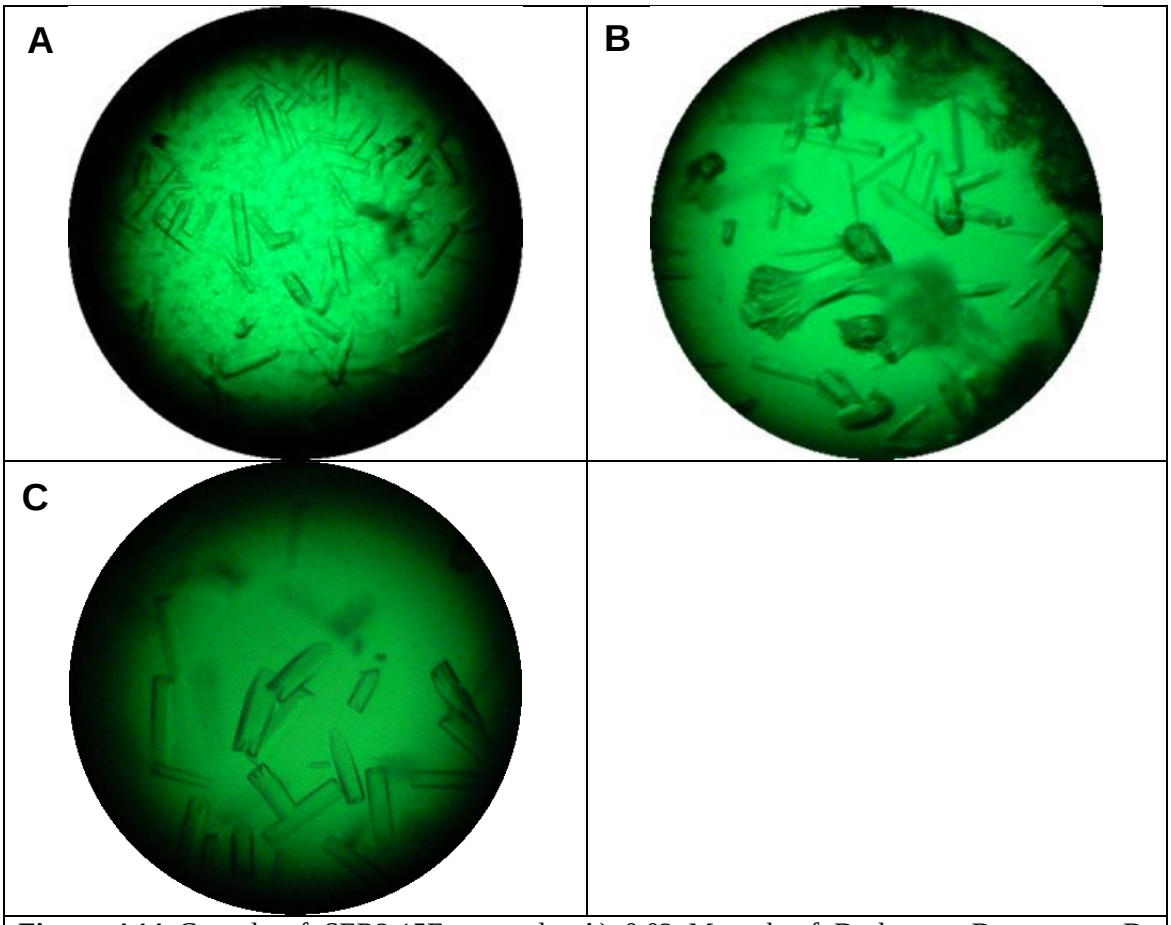


Figure 4.14 Crystals of SER2-15Ea crystals. A) 0.02 M each of D-glucose, D-mannose, D-galactose, L-fucose, 0.1 M MES/Imidazole pH 6.5, 20% Glycerol, 10% PEG4000. B) 0.1 M Sodium formate, 0.1 M MES/Imidazole pH 6.5, 20% Ethylene glycol, 8% PEG8000. C) 0.07 M D-galactose, 0.1 M MES/Imidazole pH 6.5, 20% Glycerol, 10% PEG8000.

4.6 Analysis of SER2-15Ea crystals

Diffraction datasets of SER2-15Ea crystals were collected at beamlines ID14-4 and ID23-1 of the ESRF synchrotron, Grenoble, France. Four datasets were collected and the reflections were indexed and integrated either by using iMosflm or XDS, and scaled using Scala.

The identification of the correct space group proved to be challenging. For most datasets iMosflm suggested a trigonal primitive lattice with unit cell parameters a = b = 208.23 Å, c =207.74 Å, $\alpha = \beta = 90$ °, $\gamma = 120$ °, but examination of the indexed and merged MTZ file with Pointless indicates a distinctly different orthorhombic system of with a C-centred unit cell with unit cell parameters a = 209.60 Å, b = 363.40 Å, c =212.5 Å, $\alpha = \beta = \gamma = 90$ ° as found in the space groups C222 and $C222_1$. If a crystallographic symmetry is to be hypothesised for a phytochrome dimer as it has been observed for many phytochrome structures to date, a two-fold rotational symmetry will be included somewhere in the space group symbol. Interestingly, the analysis with Pointless of a dataset 0-4 collected in September 2011 returned $P6_322$, which is in close agreement with P622 as suggested by XDS. Combining this finding with the hypotheses that two molecules in a phytochrome dimer in the crystal are related by crystallographic two-fold rotational symmetry and that the observed crystal morphology is that of hexagonal rods, the newly suggested P622 or variations like $P6_322$ were assumed to be the correct space group.

Data processing of datasets of SER2-15Ea crystals was not straight forward. Manually inspecting the positional agreement between the recorded spots and the spots predicted by iMosflm was difficult due to poor quality of the diffraction patterns. High mosaicity of the crystal leads to virtually all recorded reflections to be predicted as partial reflections, regardless of the choice of the candidate space group. The output statistics from Scala after processing in space group P622 showed that the diffraction quality of the SER2-15Ea crystals was poor (Fig. 4.15A). Furthermore, the presence of twinning was identified by L-test, which confirms the morphological indication of twinning, which indicates that a lower symmetry space group such as P6 or P3 is correct (Fig. 4.15B).

Between 4 to 6 and 18 to 22 molecules per asymmetric units were suggested for space groups $P6_322$ and P3 by Matthews software, a cell content analysis program. Two

fragments (PAS-GAF and PHY domains) of the refined structure of SER26-BV were used as search models to find the solution for the dataset. No solution for the phase problem could be found, however, by molecular replacement. The possible reasons could be any of the following factors or combinations thereof; a relatively low signal to noise ratio of the reflections across the whole resolution range, low resolution of the diffraction images and conformational differences between the structures of the search models and the structure of SER2-15Ea in *crystallo*. No further substantial progress could be made for data analysis and optimisation of crystallisation based on SER2-15Ea in the time scale of the present study, and further modification to the construct was deemed necessary to obtain better crystals.

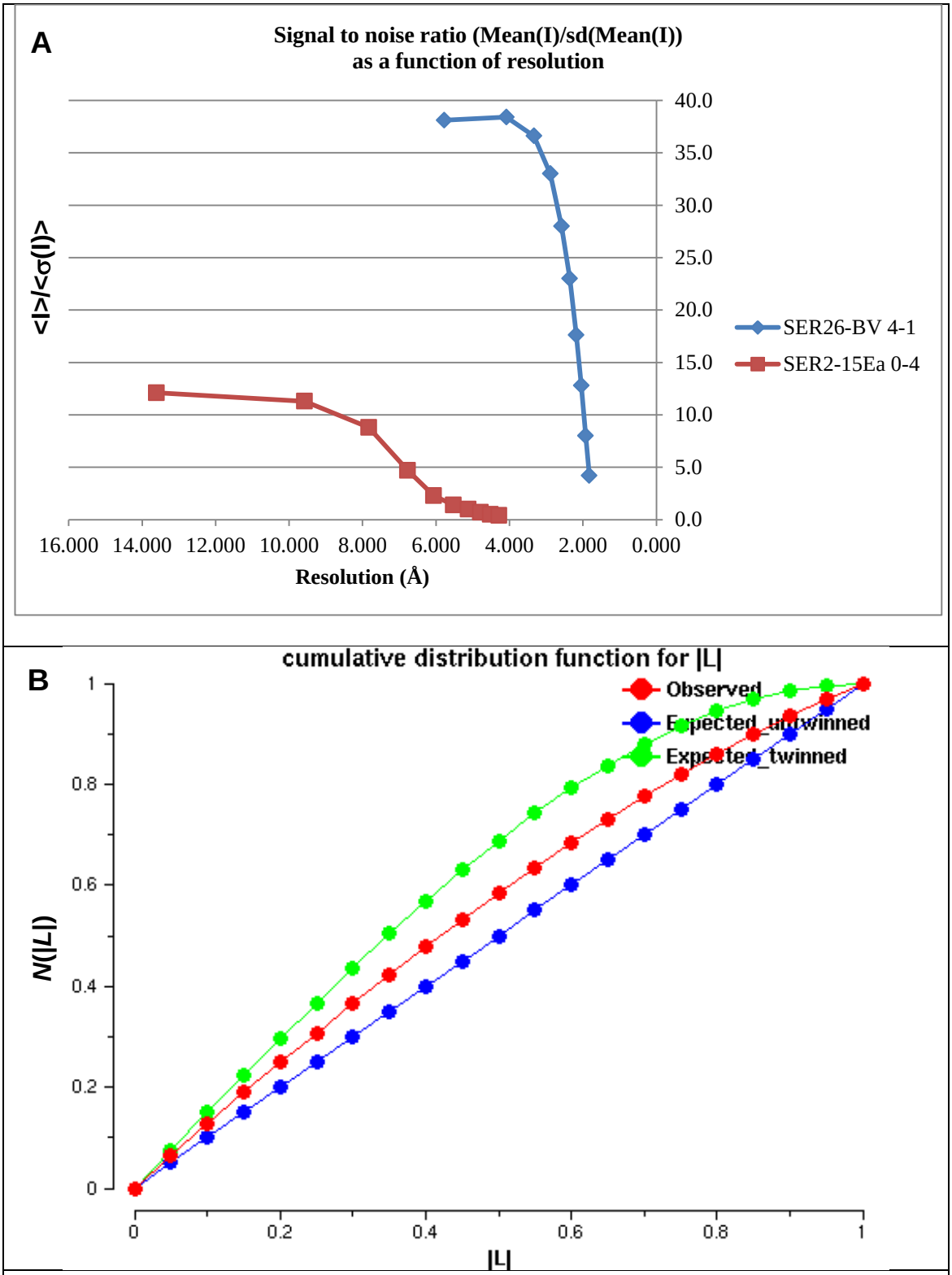


Figure 4.15 Analyses of the quality of a dataset 0-4 from a crystal of SER2-15Ea. A) Signal to noise ratio as a function of resolution is plotted for the dataset 0-4 (SER2-15Ea) and the dataset 4-1 (SER26-BV). At any given resolution, the value of Mean (I)/sd(Mean(I)) of SER2-15Ea is lower compared to that of SER26-BV, which indicates lower diffraction quality of the SER2-15Ea crystal. I = intensity. B) L-test for twinning on the dataset 0-4 collected from a SER2-15Ea crystal.

4.7 Multiple cluster surface entropy reduction mutants

Screening 15Ea adduct of various SER mutants resulted in crystallisation of SER2-15Ea. Success in obtaining crystal *de novo* is presumed to be a result of new crystal contact formation, though whether the SER2 mutation cluster directly participates in such contacts has not been explicitly shown. Structural determination of SER2-15Ea has not been successful due to poor diffraction quality of the crystals, and resulting difficulties in data processing. Additional efforts to improve the crystals such as modification of the chemical and physical condition of the crystal growth and streak/micro seeding did not result in marked improvement.

It has been shown in the present study that SER mutant proteins with similar number of mutations have markedly different crystallisation propensities, and a similar observation has been reported (Derewenda, 2011). Therefore SER23 and SER236, the constructs with the last two remaining combinations of SER mutations were produced in the hope to obtaining crystals of higher quality for the 15Ea adduct.

4.7.1 Generation of multiple cluster surface entropy reduction mutants

The constructs of SER23 and SER236 were produced by applying SER2 mutation to the constructs bearing SER3 and SER36 mutations, respectively. The SER3 construct and the SER36 construct were subjected to QuikChange mutagenesis with SER2 F and SER2 R primers (Section 2.2.2.9). Candidate plasmids were purified from overnight cultures inoculated with single colonies of *E. coli* XL1-Blue cells transformed with mutagenesis reaction mixture.

4.7.2 Verification of the plasmid constructs

An expression trial was carried out in order to confirm the successful mutagenesis. Due to the time constraint, tests for confirming correct nucleotide sequence more directly were not performed. Successful mutagenesis was judged as likely, however, from drastically altered level of soluble expression as described in Section 4.7.3.

4.7.2.1 Test expression

Expression of the protein by SER23 and SER236 construct under an inducing condition was tested in small scale test expression (Section 2.2.3.6.1). Three candidate plasmids were tested for SER23 and SER236, and they were used to transform BL21(DE3) cells.

Cells bearing candidate plasmids each were used to inoculate 5 ml of LB-Amp media, and expression of the gene of the target protein was detected by SDS-PAGE (Fig. 4.16). Overproduction of a protein with similar molecular weight as WT M15s protein was found in all cases excepting the cells bearing SER23-3 and SER236-1. In all subsequent tests, SER23-1 and SER236-2 were used.

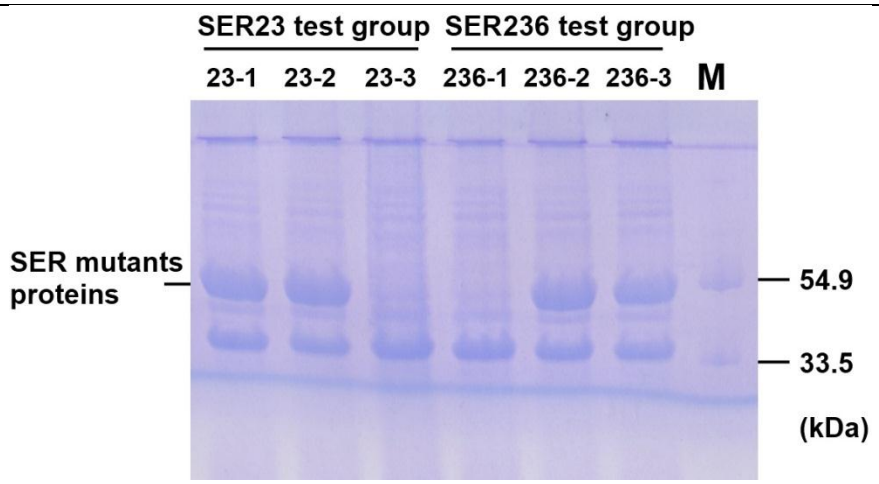


Figure 4.16 SDS-PAGE gel showing a small scale expression trial of SER23 and SER236 proteins. Successful expression resulted in proteins with the expected molecular weight for candidate plasmid number 1 and 2 of SER23, and number 2 and 3 of SER236. Intense bands with coincidentally similar molecular weight to M20s protein originate from *E. coli* insoluble fraction, and are not degradation products of overproduced proteins. M, SN2 molecular weight marker.

4.7.3 Expression and purification of multiple cluster SER mutant proteins

Large scale production of SER23 and SER236 proteins were carried out using 2100 ml of LB-amp media following the same protocol as used for the WT M15s protein (Section 2.2.3.6.2). Following 18 hours of induction, 8.58 g and 7.07 g of cells were collected for SER23 and SER236, respectively. Both proteins were purified by affinity purification as described in Section 2.2.3.7.2. Preliminary quantification of the purified protein by absorption spectroscopy indicated the total yields are 19.575 mg and 17.304 mg for SER23 and SER236, respectively. These values were gross overestimates as shown by SDS-PAGE which indicated the low proportion of target protein in the eluate from the affinity chromatography (Fig. 4.17).

The high level of contamination was also confirmed during IEX. Affinity purified SER23 was assembled with 15Ea following the same protocol as for the WT M15s-15Ea. What was initially assumed to be 19.575 mg of pure SER23 apoprotein was used for the assembly, and the sample was subjected to purification with IEX following the

same protocol as for other 15Ea adducts (Section 2.2.3.7.4). The elution profile was distinctly different from those of other 15Ea adducts with hitherto uniform pattern (Fig. 4.18). Absorption spectra of the fractions showed the peak with elution volume of 102.75 ml to contain SER23-15Ea holoprotein. SAR value of the pooled fractions containing SER23-15Ea (E2 – E6) was 0.25 which is lower than ≈ 1.00 of M15s-15Ea in the same stage of purification (Fig. 4.8). Lower SAR value is attributed to persisting contamination as shown by SDS-PAGE (Fig. 4.17). The revised estimated yield of the SER23 protein is 0.35 mg, according to the area of the SER23-containing peak relative to the total area in IEX.

The reason for a high proportion of contamination in the affinity purified samples of SER23 and SER236 was attributed to their extremely low yield of soluble protein from the cells. Affinity purification is a binding technique where the impure starting sample contains proteins with varying degree of affinity to the complementary binding substance of the stationary phase. Given finite binding capacity of the column, yield and purity of the eluted target protein dynamically varies with the abundance of target protein in the starting material, because the target protein competes for binding sites with the contaminants, albeit the former usually have higher affinity to the resin. If the target protein is not present at high enough proportion to saturate the binding capacity of the column, the remaining biding site will be occupied by contaminants. Possible reasons for the drastically reduced yields of SER23 and SER236 is discussed in Section 5.2. No further optimisation to increase the soluble yield of SER23 and SER236 was carried out.

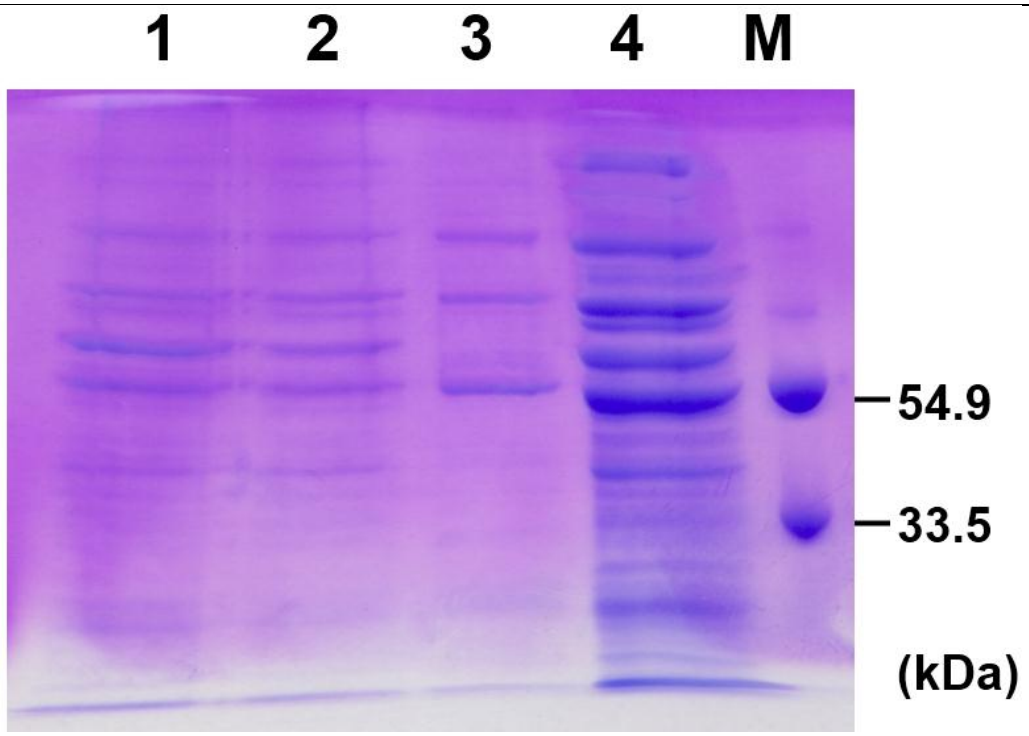


Figure 4.17 SDS-PAGE gel showing purified SER23 and SER236 proteins. Lane 1,Elution from affinity purification of SER23; Lane 2, Elution from affinity purification of SER236; Lane 3, Pooled fractions (E2-E6) from IEX containing SER23-15Ea holoprotein; Lane 4, Elution from affinity purification of SER236 at higher concentration; M, SN2 molecular weight marker. The band thought to correspond to SER23 in lane 1 and SER236 in lane 4 at 54.9 kDa constitutes 9.76% and 20.60% of the total band intensity, respectively. Following IEX purification, the relative proportion of SER23-15Ea protein rose to 28.34%. Band intensity calculation performed by calculated by ImageJ (Abramoff et al., 2004).

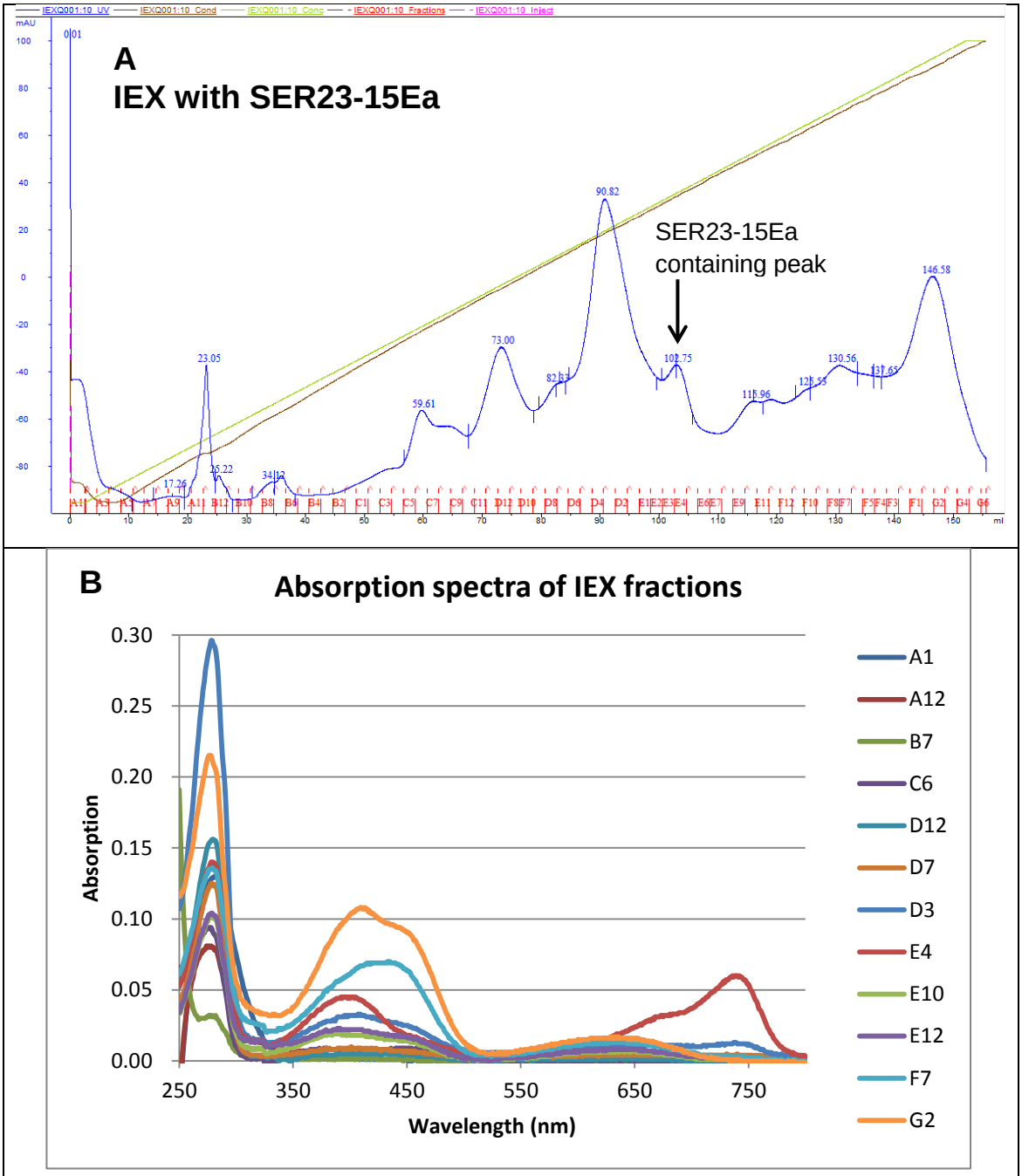


Figure 4.18 Purification of SER23-15Ea with IEX. A) IEX elution profile of SER23-15Ea. Peaks belonging to contaminations are visible. The blue, brown, and green lines indicate the absorption (mAU), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively. B) UV-vis absorption spectra of the eluted peaks from IEX. Peak with elution volume at 102.75 ml is characterised by higher SAR compared to the rest, indicating the presence of SER23-15Ea.

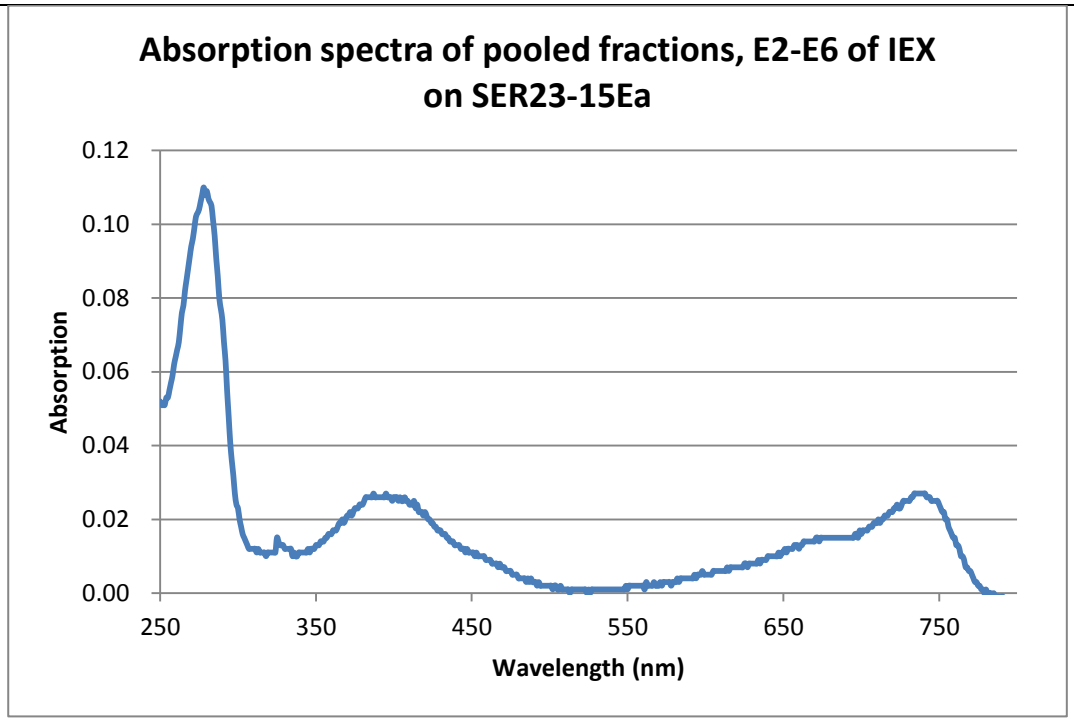


Figure 4.19 A UV-vis absorption spectrum of the pooled fractions containing SER23-15Ea holoprotein from IEX. Fractions from E2 to E6 that contain SER23-15Ea were pooled. SAR is 0.25, lower than that of WT M15-15Ea which is ≈ 1.00 and this was attributed to relatively low purity of SER23-15Ea holoprotein, and is confirmed in lane 4 of an SDS-PAGE gel in Fig. 4.17.

4.8 Discussion and future works

Following an extensive screening of the 15Ea adduct of the WT Agp1 by our collaborative groups in Berlin and London, the latter of which is described in this thesis, did not result in identification of a crystallisation condition. Screening 15Ea adducts of several SER mutants with much fewer crystallisation conditions, however, led to obtaining crystals of SER2-15Ea, the first phytochrome crystal purely in the form which represents the photoconverted state. Structural determination was not possible due to the low quality of collected diffraction images, and attempts to extend screening by producing SER23 and SER236 mutants failed due to low yields. It is plausible, however, for SER2-15Ea to serve as a basis for carrying out further modifications to the construct in the hope to obtain better crystals. Possible modifications include a conventional optimisation route of removing the C-terminal hexahistidine tag and the preceding "LE" motif derived from the XhoI restriction site. Additionally, several amino acids with low-conformational entropy are found at the C-terminus (501 KTEE 504), and removing these residues may improve the crystallisation propensity of the protein. Furthermore, introducing an Arg residue conserved in many phytochromes, a residue believed to play a key role in mediating parallel dimer formation in phytochromes (RpBphP3 Arg 150, the equivalent residue in Agp1 is Thr 131) may improve the crystal formation, because a correlation between dimerisation and the crystallisation propensity has been suggested (Bellini and Papiz, 2012a). Introduction of the dimer-mediating Arg will alter the K_D of dimer formation by Agp1, but given the wide presence of this Arg residue in other phytochromes, the structure of Agp1 with the Thr 131 Arg mutation will still be a good representative for classical bacteriophytochromes in the Pfr-form. An alternative approach in gaining functional insights into phytochrome is to carry out structural analysis on a bathy-phytochrome in its ground state (Pfr) and its adduct with 15Za, another locked chromophore which mimics the configuration of BV in the Pr-form (Inomata et al., 2006).

Should the crystal structure of Agp1 in the Pfr-form be solved in the future, the information regarding its inter-domain orientation must be interpreted carefully. It was shown in Section 3.10.2 that alternative crystal packing causes a difference in inter-domain orientation between the WT and the SER26 mutant Agp1 structures, and a similar effect is also found in Cph1 (Mailliet et al., 2011) and PaBphP (Yang et al., 2009). Since the crystal packings differ for each crystal form, it may be harder to

deduce the mechanism of signal transduction from crystal structures of a phytochrome in the Pr- and the Pfr-form than initially expected. For this reason, the full-length structure of a phytochrome in combination with results obtained by complementary techniques such as electron paramagnetic resonance (EPR) spectroscopy in combination with spin labelling may better shed light on the mechanism of phytochrome in the future.

Chapter 5: Assessment of the effects of the surface entropy reduction strategy

5.1 Overview

Introducing mutations involving charge alteration could result in destabilisation or unfolding of the protein. Concerns about a decrease in protein solubility and stability by SER mutation have been raised in the past. Four mutants of RhoGDI were reported to have lower solubility than the WT protein. The mutation comprised replacing the SER target residues (Glutamate, lysing or glutamine) with alanine, threonine, or tyrosine (Derewenda, 2011). The type of mutations and the magnitude of the decrease in solubility were overall consistent with a systematic study on the effect of the type of mutation on the solubility of ribonuclease (Trevino et al., 2007). A decrease in protein stability was observed in most of single and double SER mutants of RhoGDI, albeit the E87A mutant exhibited higher stability than E87D, therefore protein stability appears to be not directly correlated with the charge of the amino acid on the surface of the protein (Mateja et al., 2002).

Human RhoGDI is the most extensively characterised protein with regards to the effect of SER mutations on crystallisation (Cooper et al., 2007, Keep et al., 1997, Longenecker et al., 2001, Mateja et al., 2002). The effect of SER mutation on the protein conformation was tested by carrying out structural superpositioning between the structures of the WT and SER mutants, and also between different SER mutants (Table 5.1). Overall no major conformational difference was found between the WT proteins and SER mutants, and between one SER mutant to another, hence a SER mutant structure may be regarded as a good representative of that of the WT.

Human vaccinia-related kinase 1 protein is an extreme case where crystallographic structural determination was only possible following extensive SER mutations (PDB entry 3OP5, Allerston et al., to be published). Eleven SER mutations were introduced across four surface patches, which accounts for 3.02% of the primary structure. SER mutation is introduced at a more moderate degree in the range from 0.40% for SER2 (2 mutations / 504 residues) to 1.39% for SER236 (7 mutations / 504 residues) in the present study, therefore a less adverse effect, if any, is expected. In the following sections, the effects of SER mutations of Agp1 on biochemical properties and crystallisation propensity are discussed.

	1FST	1FSO	1FT0	1FT3	1KMT	2JHS	2JHT	2BXW	2JI0	2JHU	2JHV	2JHW	2JHX	2JHY	2JHZ
1RHO	0.525	0.571	0.721	0.588	0.517	0.410	0.533	0.437	0.512	0.623	0.619	0.502	0.522	0.435	0.539
1FST		0.495	0.722	0.595	0.417	0.388	0.436	0.512	0.494	0.587	0.598	0.542	0.474	0.409	0.506
1FSO			0.669	0.623	0.382	0.394	0.509	0.342	0.619	0.553	0.570	0.654	0.490	0.394	0.505
1FT0				0.282	0.600	0.460	0.859	0.729	0.571	0.402	0.415	0.612	0.476	0.478	0.476
1FT3					0.481	0.458	0.698	0.608	0.551	0.359	0.376	0.513	0.390	0.469	0.406
1KMT						0.334	0.514	0.403	0.498	0.407	0.433	0.502	0.314	0.379	0.343
2JHS							0.349	0.288	0.428	0.482	0.494	0.476	0.390	0.073	0.434
2JHT								0.421	0.473	0.685	0.689	0.582	0.523	0.356	0.592
2BXW									0.474	0.617	0.612	0.524	0.505	0.284	0.541
2JI0										0.470	0.603	0.569	0.484	0.445	0.517
2JHU											0.153	0.569	0.198	0.517	0.182
2JHV												0.557	0.241	0.526	0.216
2JHW													0.502	0.478	0.521
2JHX														0.412	0.130
2JHY															0.463

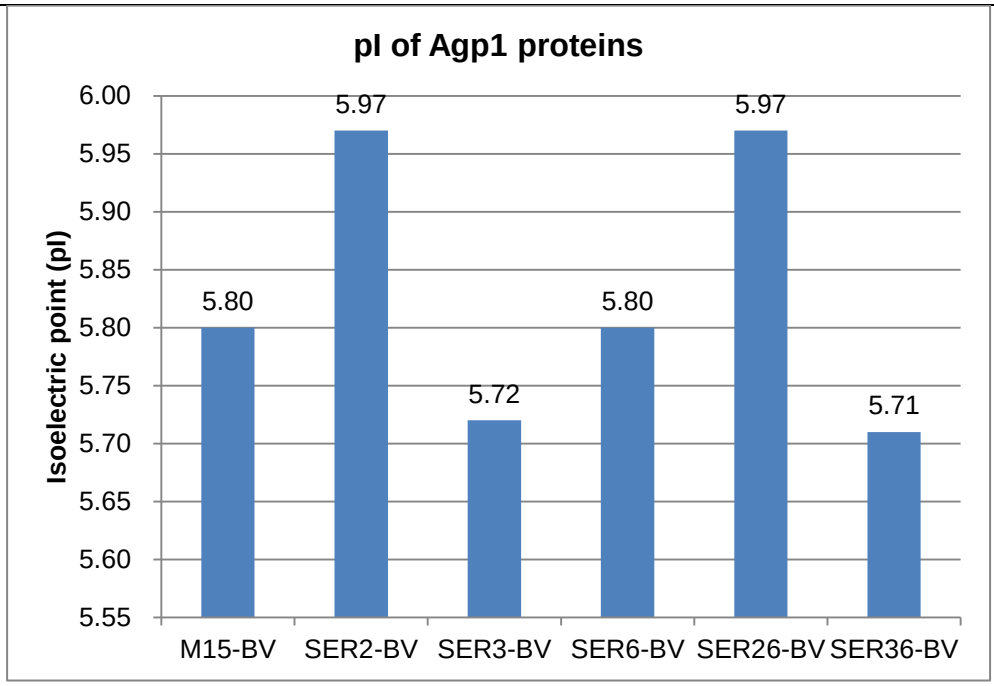
Table 5.1 Structural deviation between the WT and SER mutant, expressed as RMSD in Å. Chain A was isolated and used for structural alignment, if the initial attempted resulted in aberrant result and the asymmetric unit contains multiple chains. Structural alignment was performed with *super* command in PyMol version 1.3. All structural alignment resulted in RMSD value of equal to or less than 0.859 Å, with the average value being 0.481Å, indicating a conserved fold regardless of the introduction of different mutations or different crystal packing. 1RHO (WT) (Keep et al., 1997); 1FST (K135/138/141A), 1FT0 (K113A), 1FT3 (K141A) (Longenecker et al., 2001); 1KMT (E154A/155A) (Mateja et al., 2002); 2JHS (K135/138/141H), 2JHT (K135/138/141T), 2BXW (K135/138/141Y), 2JI0 (K138/141Y), 2JHU (E154/155A), 2JHV (E154/155A), 2JHW (E155/157A), 2JHX (E155/157H), 2JHY (E155/157H), 2JHZ (E155/157S) (Cooper et al., 2007).

5.2 Effect of surface entropy reduction mutation on the solubility of the proteins

The decrease in solubility caused by SER mutations was expected since the corresponding structural changes involve replacing the charged residues (Glu and/or Lys) with the non-polar Ala (Derewenda, 2011). The yield of soluble protein from affinity purification may be regarded as an indication of solubility of the protein the cytoplasm of *E. coli*, but the yield is also affected by multitudes of other factors. To this end, an objective study was undertaken to compare the solubility of the WT and mutants of Agp1 proteins. Since Agp1 proteins have high solubility in the SEC buffer (Section 2.2.3.7.6), the relative solubility of the proteins were compared under the presence of 1.2 M $(NH_4)_2SO_4$ to emulate the method described in Trevino et al. (2007) as described in Section 2.2.4.6. The result is shown in Figure 5.1. SER2-BV and SER3-BV have lower solubility than that of the WT M15-BV, with SER3-BV being marginally less soluble than SER2-BV. SER6-BV has a higher solubility than the WT M15-BV. The presence of the SER6 mutation increases the solubility even in the presence of other SER mutation clusters, SER2 and SER3. Solubility of SER26-BV and SER36-BV are higher than that of SER2-BV and SER3-BV, respectively. Relative solubility of SER26-BV and SER36-BV are the same as when those proteins are without the SER6 mutation, with SER26-BV being more soluble than SER36-BV.

The pI is known to influence the solubility of a protein (Tan et al., 1998, Wu et al., 2010). When the pH of the buffer is equal to the pI of a protein, the net charge of the protein is zero. The attractive electrostatic forces due to the dipole moment of the protein promote non-specific self-aggregation between proteins. Conversely, charged protein will resist aggregation by the repulsive electrostatic force. According to this rationale, a mutant protein with a lower pI than Agp1 M15s (pI 5.80) should have higher solubility than the WT protein. This will lead to a prediction that a E→A mutation will impact the solubility while K→A mutation will promote solubility by the virtue of lowering the pI. Correlation between the calculated pI of proteins and their solubility is difficult to establish, however. According to the above rationale, the SER3 protein (E96A/K97A/K98A) should feature a higher solubility than the SER6 protein (E336A/K337A) by the virtue of an additional K→A mutation, but SER6-BV had a higher solubility even than the WT M15-BV. Solubility of SER23-BV and SER236-BV

could not be tested due to being unable to obtain sufficient amount of purified apoprotein within the duration of the present study. This could imply that the simultaneous presence of SER2 and SER3 mutations, the two mutations that reduces the solubility of the protein, synergistically reduced solubility of the proteins to significantly reduce the proportion of soluble protein in the *E. coli* cytoplasm. The combination of the two sets of mutations is hypothesised to synergistically create a high concentration of hydrophobic residues on the protein surface which impacts solubility more than when SER mutation clusters are introduced singly in a limited region. As of now little is understood about how a local effect like the interaction between an amino acid side chain and the solvent determines a global property such as the solubility of a protein.



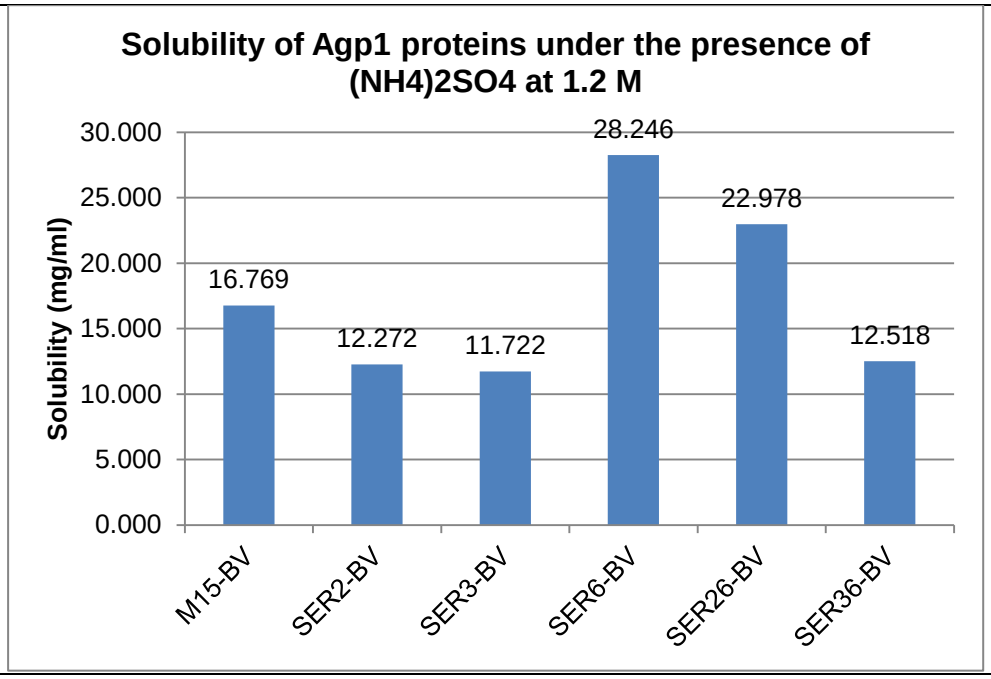


Figure 5.1 The calculated pI and the solubility of BV adducts of Agp1 proteins. A) The pI of Agp1 protein variants calculated by Protein Calculator (http://www.scripps.edu/~cdputnam/protcalc.html). B) Solubility of Agp1 protein variants as quantified spectrophotometrically under the presence of 1.2 M of $(NH_4)_2SO_4$. Introduction of single cluster SER mutations reduces the yields of apoprotein in varying degrees depending on the location of the mutation site.

5.3 Biochemical verification of the SER mutants

5.3.1 UV-vis absorption spectroscopy

The ability to undergo photoconversion is a key aspect of phytochrome function. In order to verify their functionality, BV adducts of the SER mutants were tested for the ability to form the Pfr-form in response to red light.

15Ea adducts of Agp1 representing the Pfr-form were omitted from the test, as they do not undergo photoconversion. Absorption spectra of 15Ea adducts of SER mutants are shown in Section 4.5.1, and they are indistinguishable from the WT holoprotein.

Absorption spectra of BV adducts of M15s as WT, SER2, SER3, SER6, SER26, and SER36 at about 0.1 mg/ml were taken in the Pr-form, the dark adapted from. The ability of SER mutants to undergo photoconversion similarly to the WT holoprotein was confirmed by illuminating the sample with a red laser (650 +/- 20 nm) until no further change in absorption spectrum was observed. In all cases, BV-adducts of all SER mutants produced absorption spectra indistinguishable from the WT in both Pr (Fig. 5.2) and Pfr-forms (Section4.5.1), except for SER36-BV. The slightly different spectrum of SER36-BV in the Pfr-form may be attributed to the heterogeneity of the sample. The experiment proves that the spectrophotometric properties are not affected in most SER mutants.

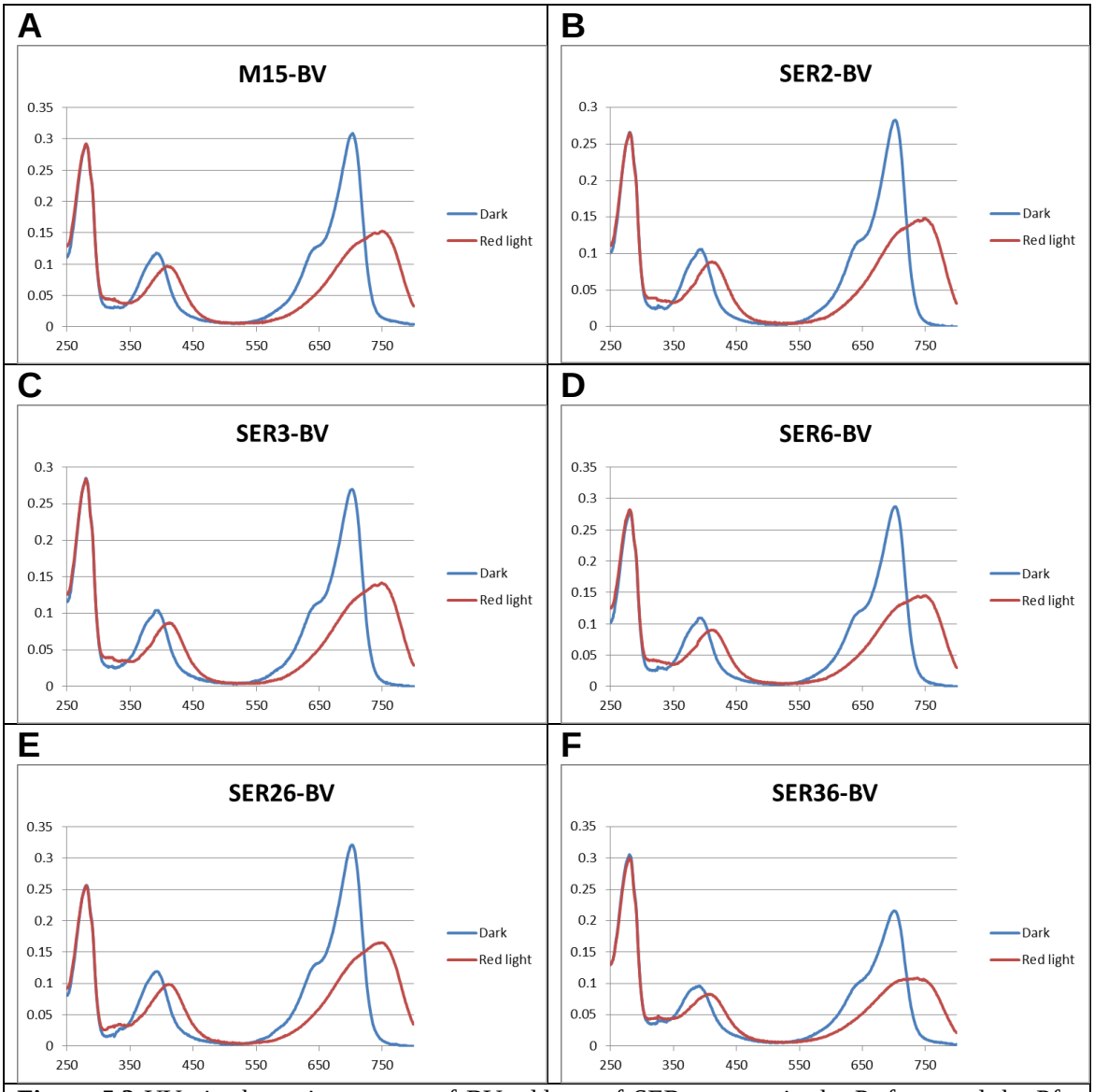


Figure 5.2 UV-vis absorption spectra of BV adducts of SER mutants in the Pr-form and the Pfr-form. Dark adapted protein samples in the Pr-form were measured first (blue line), then the protein solution in the cuvette was exposed to red laser light until no further changes in the spectra was observed (red line).

5.3.2 Size exclusion chromatography

The tertiary and quaternary structures of the SER mutants were probed with SEC. Protein concentrations of the holoproteins (M15s-BV, SER2-BV, SER3-BV, SER6-BV, SER26-BV, SER36-BV, M15s-15Ea, and SER2-15Ea) were adjusted to 1 mg/ml. 100 µl of the samples were injected into a 250 µl sample loop, and subjected to analytical SEC on a Superdex200 column with SEC buffer (Section 2.2.3.7.6). In all cases the elution volumes of all SER mutants were similar to that of the WT, therefore the SER mutations did not alter the tertiary and quaternary structure significantly (Fig. 5.3). Additionally, the SEC elution volume of the 15Ea adducts of the SER mutants were smaller than those of the BV adducts. This form-dependence in SEC elution volume is in accordance with previous publications for the WT protein (Inomata et al., 2005, Noack et al., 2007), and confirms that the SER mutations do not cause major structural changes.

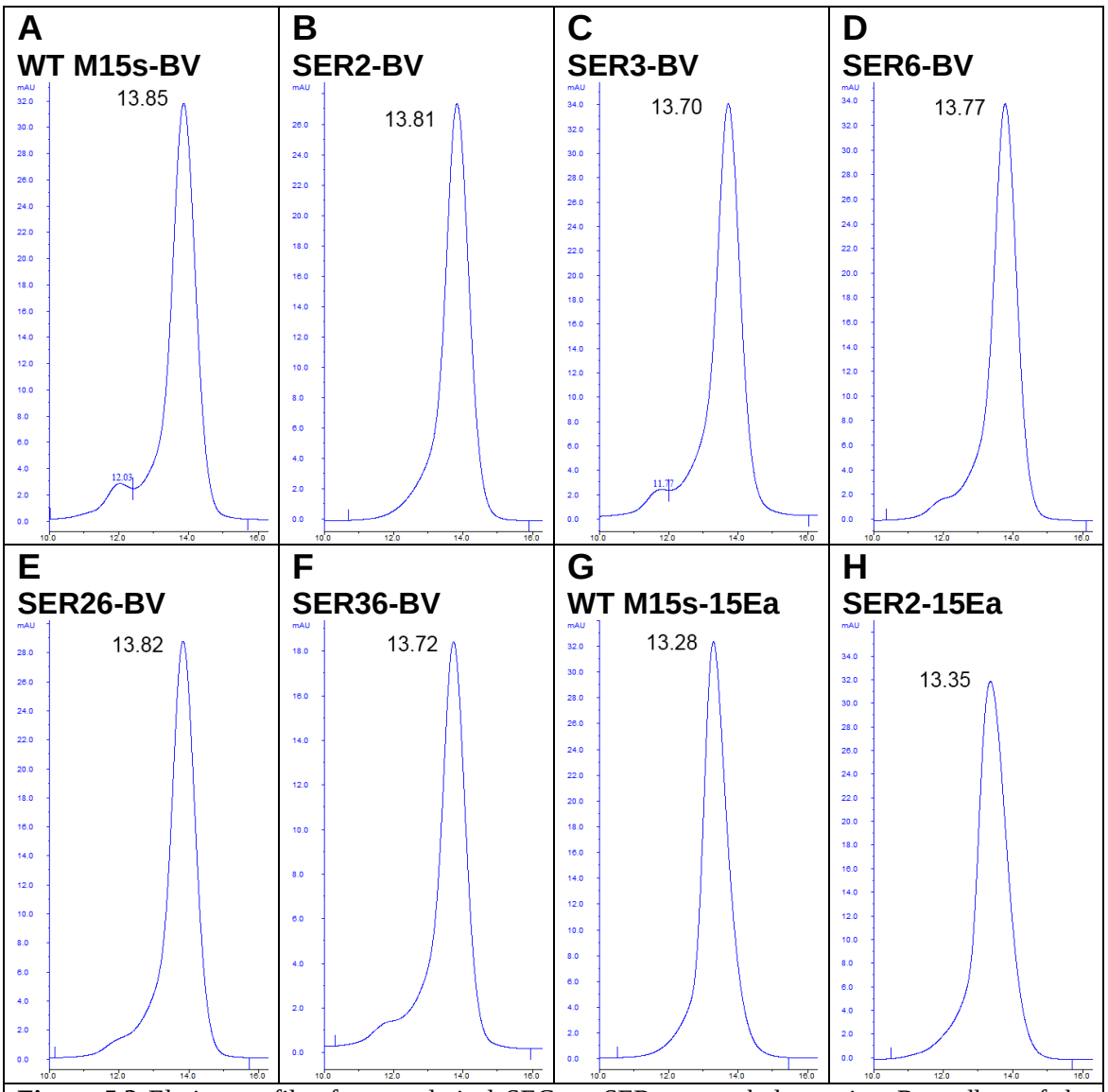


Figure 5.3 Elution profiles from analytical SEC on SER mutant holoproteins. Regardless of the protein, the elution volume given in ml in each panel is influenced by the chromophore (BV or 15Ea), suggesting that SER mutation does not alter the tertiary and quaternary structures of the holoprotein. Chromatograms were obtained by monitoring absorption at 280 nm. A) M15s-BV; B) SER2-BV; C) SER3-BV; D) SER6-BV; E) SER26-BV; F) SER36-BV; G) M15s-15Ea; H) SER2-15Ea.

5.3.3 Circular dichroism spectroscopy of SER holoproteins

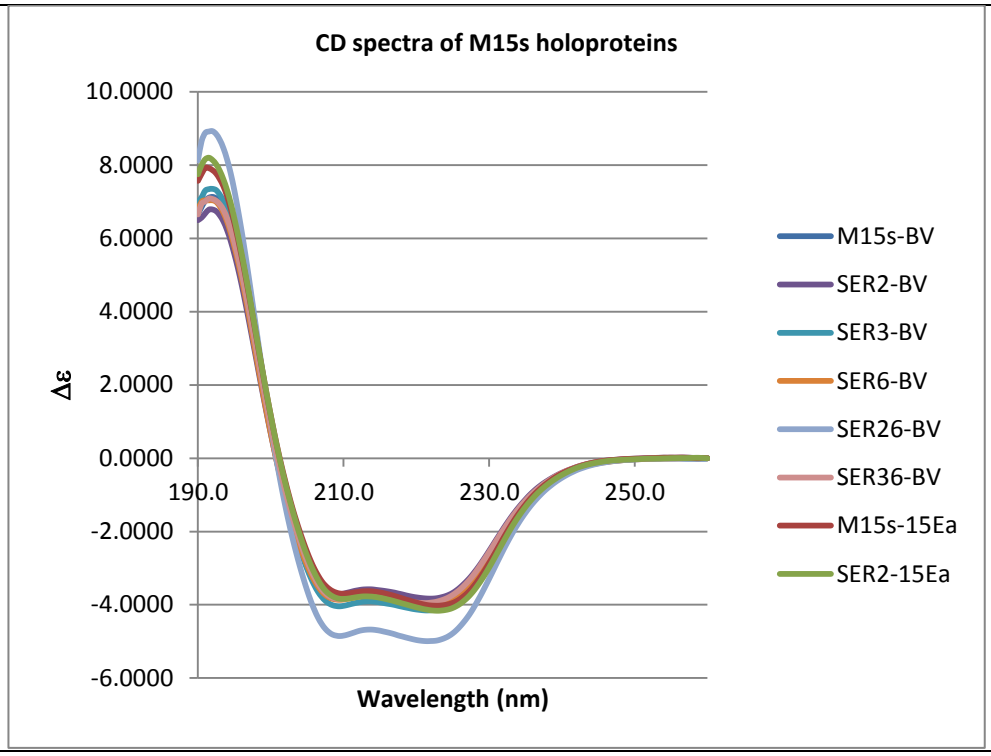
The effect of the SER mutations on the secondary structures of the holoproteins was investigated by circular dichroism (CD) spectroscopy in the far-UV range (190 – 260 nm). Concentrated holoprotein samples at typically 20 mg/ml in IEX start buffer were diluted with CD buffer to 0.1 mg/ml. The absorption at 280nm was taken immediately prior to the CD measurement, in order to calculate the molar circular dichroism ($\Delta\varepsilon$) afterwards.

The results indicate that the proportions of secondary structure elements are comparable within the holoproteins with the same chromophore, thus indicating that SER mutations do not alter the secondary structure content of the protein (Fig. 5.4). The CD-spectrum of SER26-BV stands out from other BV-adducts, but this may be attributed to an experimental error during the quantification of the protein concentration. The secondary structure content calculated from the structure by DSSP is more consistent with the values obtained for other BV-adducts, as calculated from CD spectra.

There are slight differences in the CD-spectra of adducts of BV and 15Ea. Peaks at approximately 193 nm in the spectra of 15Ea adducts are higher than those of BV adducts. Additionally, the relative intensities of the two minima (approximately at 210 nm and 224 nm) in spectra of BV and 15Ea adducts are different i.e. the relative intensity of the minimum at 224 nm is higher for the 15Ea adducts. Secondary structure content calculated from these spectra reveal no obvious difference between the BV and 15Ea adducts, however.

Difference in the secondary structure content of BV and 15Ea adducts of the WT protein is discussed in Section 6.4.2.

A



B

Protein	α-helix	β-sheet	Turn	Disorder	Total
M15s-BV	0.39	0.16	0.17	0.27	0.99
SER2-BV	0.37	0.16	0.18	0.28	0.99
SER3-BV	0.40	0.16	0.17	0.27	1.00
SER6-BV	0.38	0.16	0.18	0.28	1.00
SER26-BV	0.49	0.11	0.15	0.26	1.01
SER36-BV	0.38	0.17	0.18	0.27	1.00
M15s-15Ea	0.40	0.16	0.18	0.26	1.00
SER2-15Ea	0.40	0.16	0.17	0.27	1.00

C

Protein	α-helix	Isolated β-bridge	b-strand	3-helix (3/10 helix)	H-bonded turn	Bend	Loop / Irregular
SER26-BV	0.35	0.01	0.26	0.02	0.11	0.08	0.17

Figure 5.4 CD spectra of M15s holoproteins as BV and 15Ea adducts. A) Molar circular dichroism (Δε) of several holoproteins studied in this thesis. B) Secondary structure content calculated from CD-spectra with Dichroweb. C) Secondary structure content of SER26-BV structure, calculated with DSSP (Joosten et al., 2011). The proportion of each secondary structure element is given with respect to the number of modelled residues.

5.4 Correlation between crystal contact mediating epitope and possible crystal form

It was noticed during crystallisation screening that all BV adducts of SER mutants produced positive signs of crystallisation under conditions similar to the one which leads to crystallisation of the WT M15(s)-BV. Herein, a positive sign of crystallisation is subjectively defined as green crystalline precipitate (score 5 in a scheme put forward by Benvenuti and Mangani, 2007) or better. In Section 3.7.3, SER3-BV and SER36-BV are described to crystallise under the same condition to form the same crystal form, but this condition and the crystal form are distinct from those of the WT crystal. This and other observations led to hypothesising a correlation between the presence of a particular epitope and the possible crystal form of the Agp1 proteins. To this end, crystallisation experiments were carried out. It must be noted however, that the formation of a nucleus in a crystallisation drop is a stochastic event, therefore a negative result does not necessarily prove the inability of a protein to crystallise under a given condition.

Firstly, the solution of the BV adducts of the WT and some SER mutants were adjusted to the identical concentration of 25 mg/ml, and subjected to crystallisation under the same crystallisation condition as used for the WT M15(s)-BV protein. The results indicate positive signs of crystallisation for all holoproteins, albeit with varying degree in crystal quality (Fig. 5.5). This shows the capacity of all SER mutants tested in the present study to crystallise with what is assumed to be the same form as the WT. If the structural information from the WT crystal is reliable, it would have been interesting to investigate if the residues participating in the crystal contact of the WT crystal form are also conserved in the SER mutants in terms of the structure.

A similar test was carried out, where SER3-BV, SER36-BV and WT M15s-BV were subjected to crystallisation under conditions derived from the $(NH_4)_2SO_4$-based condition identified for SER36-BV. The trial resulted in successful crystallisation of SER3-BV and SER36-BV, but not of the WT protein (Fig. 5.6). This is attributed to the common presence of the SER3 mutated epitope (E96A/K97A/K98A) to mediate a crystal contact that allows for this particular crystal form to appear under $(NH_4)_2SO_4$-based conditions.

A crystallographic trial to test if the simultaneous presence of SER2 and SER6 mutations were necessary for the crystallogenesis of SER26-BV protein in a novel form would have been interesting. Crystallisation of SER2-BV or SER6-BV with the same crystallisation condition as for SER26-BV could not be tested due to the limited time available for the present study, but an analysis using the crystal structure of SER26-BV is described in Section 5.5. The conclusion from the analysis would mean that the simultaneous presence of SER2 and SER6 mutations is necessary for the crystallogenesis of SER26-BV.

Conversely to the above observations, SER26-15Ea crystals were never observed under the condition which facilitates the growth of SER2-15Ea. This is different from the above cases since an additional SER cluster seems to prohibit the formation of crystals. The unexpected absence of SER26-15Ea crystal formation may be due to reasons other than the crystallisation propensity of the protein (e.g. sample heterogeneity). An alternative proposal is that an additional SER6 cluster might be involved in a formation of a new interface between monomers, therefore the equilibrium of thermodynamically favourable crystal packing from is slightly shifted from one form to the other, thereby inhibiting continuous crystal growth with a consistent lattice.

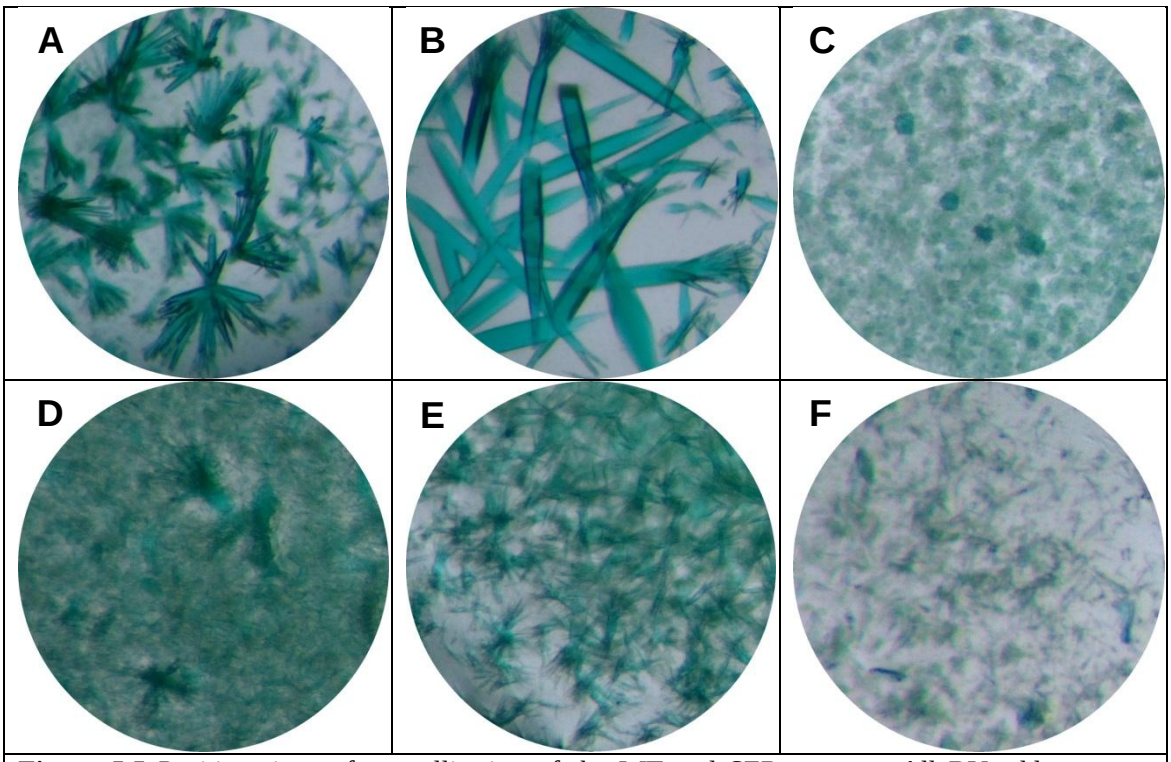


Figure 5.5 Positive signs of crystallisation of the WT and SER mutants. All BV adducts were subjected to crystallisation under the optimal crystallisation condition for the WT M15(s)-BV protein. A) M15s-BV. B) SER2-BV. C) SER3-BV. D) SER6-BV. E) SER26-BV. F) SER36-BV. In all cases, the trial resulted in crystalline precipitate or better. The crystalline precipitates are unlikely to be denatured aggregate, since the precipitates samples still retain the green colour, indicative of conservation of the tertiary structure in Agp1 holoproteins.

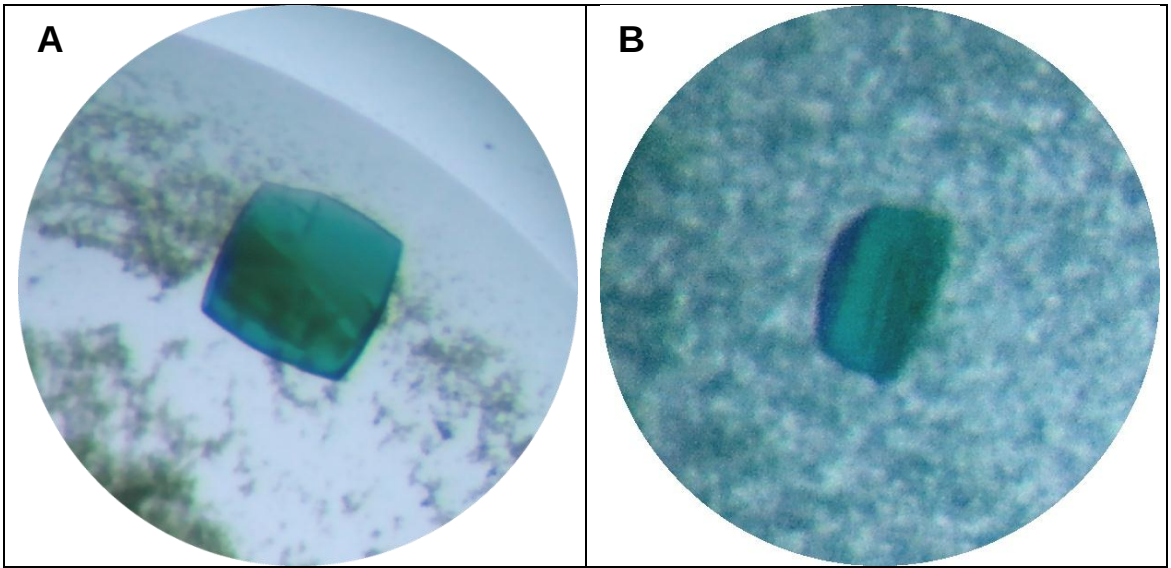


Figure 5.6 Crystals of SER3-BV and SER36-BV. A) SER3-BV and B) SER36-BV crystals. Both crystals were obtained by minor variation of the condition identified for SER36-BV as described in Section 3.7.4.

5.5 Effect of SER mutation on the crystal packing

Introduction of SER mutations resulted in crystallisation of M15-BV in alternative crystal forms, and the correlation between the presence of a particular epitope and a specific crystal form was discussed in Section 5.4. The effect of Ala substitution of Glu/Lys residues on crystallogenesis of SER26-BV was investigating by simulating simple mutagenesis using the crystal structure of SER26-BV (Fig. 5.7). The "simple mutate" function of Coot was used to revert the SER mutations (Site 2, Ala86Glu / Ala87Glu; Site 6, Ala336Glu / Ala337Lys), followed by idealization of the geometries of the modified residues. Inspection of the two reverted sites in the SER26-BV structure indicates that in either case, the original WT residues could cause electrostatic repulsion with the residues of symmetry related molecules. The closest distances between Glu 86 and Glu 87 of one molecule and Glu 489 of a symmetry related molecule are 5.5 and 3.5 Å, respectively, and the closest distance between Lys 337 of one molecule and Lys 220 of another symmetry related molecule is 4.8 Å. These observations suggest that electrostatic repulsion between equally charged residues in the case of WT M15s-BV prevented formation of the novel crystal form that was observed for SER26-BV.

The rationale behind Ala substitution on Glu/Lys residues in the surface entropy reduction strategy is to decrease the surface conformational entropy, but successful crystallogenesis of SER26-BV in a new crystal form may also be attributed to resolving electrostatic repulsion.

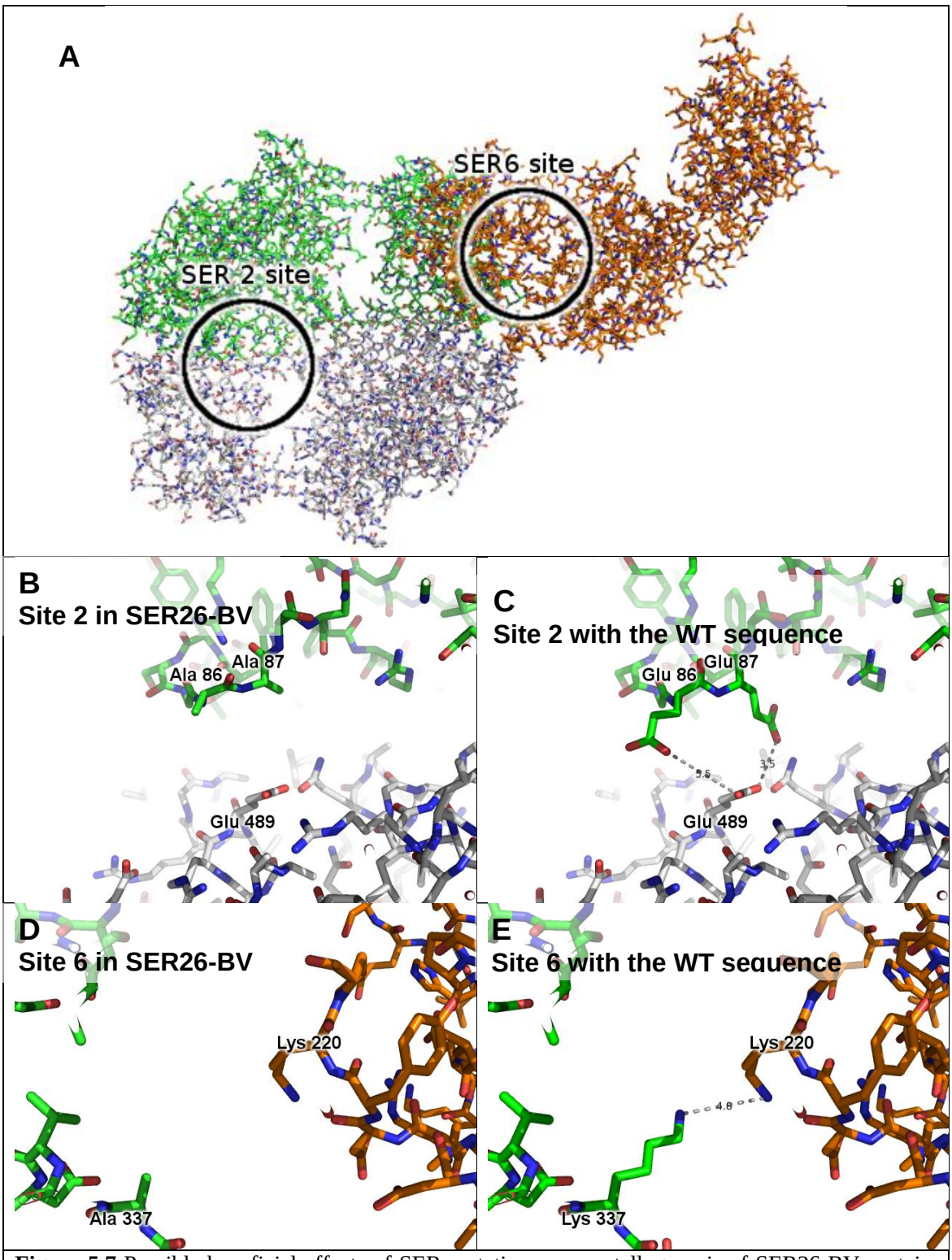


Figure 5.7 Possible beneficial effects of SER mutations on crystallogenesis of SER26-BV protein. Mutation of SER sites with the original Glu/Lys residues were simulated, followed by idealisation of the geometry of the mutated residues with Coot. A) Overview of the locations of the SER mutation sites 2 and 6 in the SER26-BV crystal structure. B) Close-up view of SER mutation site 2 (Glu86Ala/Glu87Ala), C) and of the corresponding site in the WT model. The presence of Glu residues in SER site 2 may cause electrostatic repulsion with Glu 489 of a symmetry related molecule. Short distances are indicated by broken lines. D) Close up view of SER mutation site 6 (E336A/K337A), E) and of the corresponding site in the WT model. The presence of Lys in 337th position may cause electrostatic repulsion with Lys 220 of a symmetry related molecule.

5.6 Discussion

A correlation between the presence of an epitope and the form in which Agp1 can crystallise was described (Section 5.4). It is difficult to implement this information to predict the outcome of a crystallisation experiment *ab initio*, because at present it is impossible to predict the ability of a protein to crystallise in a novel from the amino acid sequence alone (e.g. SER26-BV). Nevertheless, it may be an effective approach to carry out crystallogenesis optimisation by exploring the SER mutants with common SER mutations to the one for which a positive sign of crystallisation has been identified.

Inspection of the SER mutation site at the molecular level suggests that the successful crystallisation of SER26 in a novel crystal form is possibly due to resolving electrostatic repulsions, rather than solely to reducing the conformational entropy. This example shows that it is difficult to predict the exact effect mutations will impose on crystallisation, especially when a change in crystal form is involved. Nevertheless, surface entropy reduction and other similar techniques that directly address the "crystallisability" of the protein will play increasingly important roles in the future. It has been suggested that proteins evolved the "entropy shield" to help retain biological function by avoiding non-specific aggregation (Doye et al., 2004). Such notion is in accordance with the first of two major bottlenecks which were identified in statistical studies from structural genomics; the first of which is to obtain crystals of the target protein and the second one is to obtain useful diffraction data which leads to structure determination (Terwilliger et al., 2009, Bergfors, 2009: p313). In the present thesis, the potential of the surface entropy reduction strategy to tackle both types of bottlenecks was demonstrated for two types of holoproteins. The problem of obtaining crystals where no previous lead condition has been identified, described as the first bottleneck, was solved for the first time by using SER2 to crystallise an Agp1 holoprotein containing the locked 15Ea chromophore. While the diffraction datasets of SER2-15Ea crystals did not result in the determination of their structure, the success of growing crystals at least opens a new avenue for further modification and optimisation to be carried out on SER2-15Ea. A solution to alleviate the second bottleneck was demonstrated by changing the crystal form of M15s-BV as demonstrated with SER3-BV, SER36-BV, and SER26-BV. Retrospectively, only one of two novel crystal forms of the BV adducts resulted in successful structure determination, and this highlights the

importance of carrying out extensive crystallisation trials where the protein itself may also regarded as an experimental variable (Longenecker et al., 2001).

Chapter 6 Biochemical characterisations of Agp1 in different forms

6.1 Overview

Structural characterisation of Agp1 in the Pr-form is described in Chapter 3. While the present study has led to successful structure determination of this form, the original aim of performing an unbiased comparative study on Agp1 structures in two forms was not possible because the crystals of the 15Ea-adduct were of insufficient quality for structure determination (Chapter 4).

Furthermore, structural information on phytochrome apoproteins is currently limited. All previous attempts to crystallise the WT Agp1 M15 apoprotein failed (N. Krauss, personal communication), therefore apoproteins of SER2, SER3, and SER26 mutants were screened with JCSG+, PACT-premier, and Morpheus kits (Molecular Dimensions), but no lead condition was identified.

In this chapter, analysis on phytochromes in three forms – apoprotein, BV adduct (Pr-form) and 15Ea adduct (Pfr-like form) with multitudes of biochemical techniques are described in an attempt to characterise the conformational differences between these forms.

6.2 Ion exchange chromatography

Previously, SEC has been successfully employed to study the conformational change associated with the photoconversion of phytochromes (Inomata et al., 2005, Inomata et al., 2006, Lamparter et al., 2001). SEC separates molecules according to their Stokes radius which depends on the molecular weight and the conformation of the protein, but knowledge of only this property could lead to difficulties in differentiating the alternative forms of the same protein, even with a column with high-resolving resin with appropriate selectivity within a relevant molecular weight range. For example, no difference in elution volumes was reported for full-length Agp1 in the Pr and Pfr-forms (Inomata et al., 2005). This is because the conformational change was insufficient to cause significant difference in the mobility of each form.

IEX differentiates between proteins according to their net surface charge. It was hypothesised that IEX may be more suitable to detect the conformational change that is indiscernible by SEC if a conformational change involving a small change in Stokes radius results in an altered net surface charge. The use of IEX will not only enable to

investigate the presence and the nature of a subtle conformational change, but as a practical implication also to purify a conformationally homogeneous species.

6.2.1 M15 and M20 as apoprotein and BV adduct

The process of bilin assembly whereby phytochromes acquire the ability to photoconvert is not well characterised. Conformational differences between apoprotein and holoprotein of Agp1-M15 and Cph1 proteins have been demonstrated with SEC (Inomata et al., 2005, Lamparter et al., 2001), but the nature of the conformational change associated with the assembly process has not been well characterised. To this end, Agp1 M15 and M20 proteins were subjected to IEX using a linear gradient elution either as apoprotein or BV adduct. Differences in the elution volume between apoprotein and holoprotein were interpreted as to indicate the presence and the quality of the conformational change. Additionally, the degree of shift in elution volume between M15 and M20 will delimit the location of the conformational change, because M20 is shorter than M15 due to the lack of the PHY domain. Such interpretation is based on the hypothesis that the nature of the conformational change during the bilin uptake is similar between M15 and M20. The hypothesis is deemed plausible given the similar increase in the absorbance in the Q-band during the chromophore assembly by both M15 and M20 proteins and the structural similarity between SER26-BV (PAS, GAF and PHY domains) and DrBphP (PAS and GAF domains) (Chapter 3).

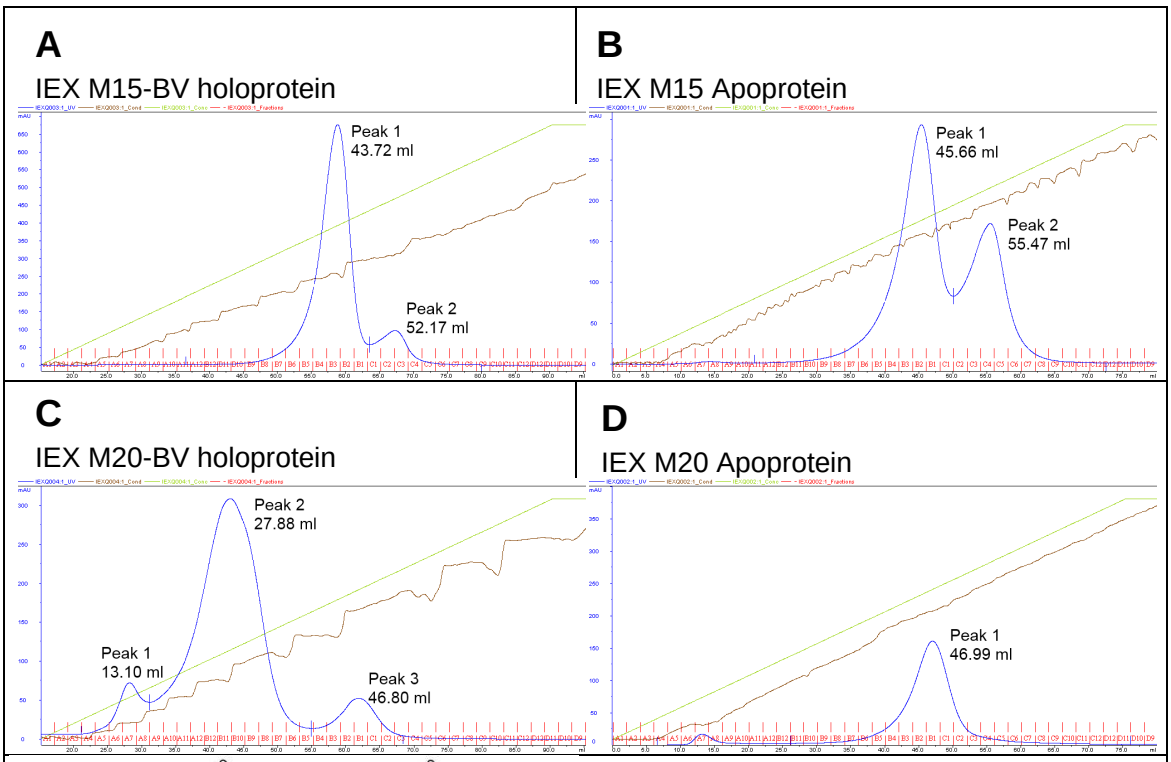
The GAF domain was hypothesised to be the major site of conformational change as it contains the bilin-binding pocket, the major site of protein-chromophore interaction. It is immediately obvious by observing the Agp1 and other phytochrome structures that the bilin is shielded from the solvent (Fig. 3.27). Similar elution volumes of holoprotein and apoprotein of Agp1-M15 protein in SEC rule out an extreme conformational difference between the two forms (Inomata et al., 2005), but clearly the apoprotein must adopt a different conformation from that of the holoprotein in order to allow for the initial uptake of the bilin into the pocket.

9000 mAU×ml (Section 2.2.3.7.4) of affinity purified M15 and M20 apoproteins were prepared either as apoprotein or BV-adduct (Section 3.4.1 and 3.2.1, respectively). M20-BV used in this experiment had not been further purified by IEX. Each protein was applied to a HiTrap Q column and eluted by a linear gradient from IEX Buffer A to

IEX Buffer B over 15 CV with a constant flow rate at 3.5 ml/min (Section 2.2.3.7.4), resulting in the chromatograms shown in Fig. 6.1.

Both M15 apoprotein and holoprotein eluted as two peaks. The earlier main peak and the second minor peak were found to consist of monomer and dimer, respectively, as shown independently by a native-PAGE and dynamic light scattering (DLS) (Fig. 6.1), as well as UV-vis absorption spectroscopy (Fig. 6.2). It is interesting that the proportion of the dimer is higher in the apoprotein, which is discussed in more detail in Section 6.2.2. The elution volumes of both monomer and dimer peaks are lower in case of M15-BV compared to the apoprotein. This indicates that the BV assembly process involves a reduction of the net negative charge on the protein surface in both monomer and dimer.

The elution profiles of M20 proteins are different from that of M15. The M20 apoprotein eluted as the dimer in a single peak only, and the equivalent peak is also present in the chromatogram of M20-BV (Fig. 6.1). The constituent of the peaks were analysed by UV-vis absorption spectroscopy (Fig. 6.2). The presence of the apoprotein peak in the chromatogram of the holoprotein indicates that the BV assembly process is inefficient in M20. M20-BV eluted in two peaks, but both peaks were found to consist of the monomeric form. This indicates that M20-BV exists predominantly in the monomeric form, whereas the apoprotein has a strong propensity to form a dimer even more effectively than M15. This is surprising because the shorter M20 has a smaller dimer interface area than M15, but its strong propensity to form a dimer may be rationalised by formation of an inter-molecular disulphide bond which is discussed in Section 6.5.1. The difference in the elution volumes of apoprotein and holoprotein is of practical importance as it enables the separation of apoprotein and holoprotein of M20, a protein whose BV assembly is inefficient. In terms of conformational difference between apoprotein and holoprotein of M20, no direct comparison could be made based on their elution volumes because the oligomeric state is different.



Sample	Molecular weight (kDa)	Polydispersity (%)
M15-BV peak 1	86	11.7
M15-BV peak 2	369	53.6
M15-apo peak 1	96	25.2
M15-apo peak 2	310	46.6
M20-BV peak 1	56	21.5
M20-BV peak 2	62	12.0
M20-BV peak 3	196	27.7
M20-apo peak 1	182	13.6

Figure 6.1 Conformational differences between Agp1-M15 and M20 in the apo- and holoprotein forms. Analytical ion exchange (IEX) chromatograms of (A) Agp1-M15 holoprotein, (B) Agp1-M20 holoprotein, (C) Agp1-M15 apoprotein, and (D) Agp1-M20 apoprotein. Values indicated next to the peaks are elution volumes from the start of the linear gradient. (E) A native-PAGE gel showing the peak-containing fractions from the IEX analyses. Based on the relative migration rates of the bands and the molecular weight of the monomer of each protein, the oligomerisation state was assigned to each band. (F) Results from DLS of the peak fractions from IEX chromatography. The measured approximate values of the molecular weight are in agreement with the relative molecular weights determined by native-PAGE. The blue, brown, and green lines indicate the absorption (mAU at 280 nm), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively.

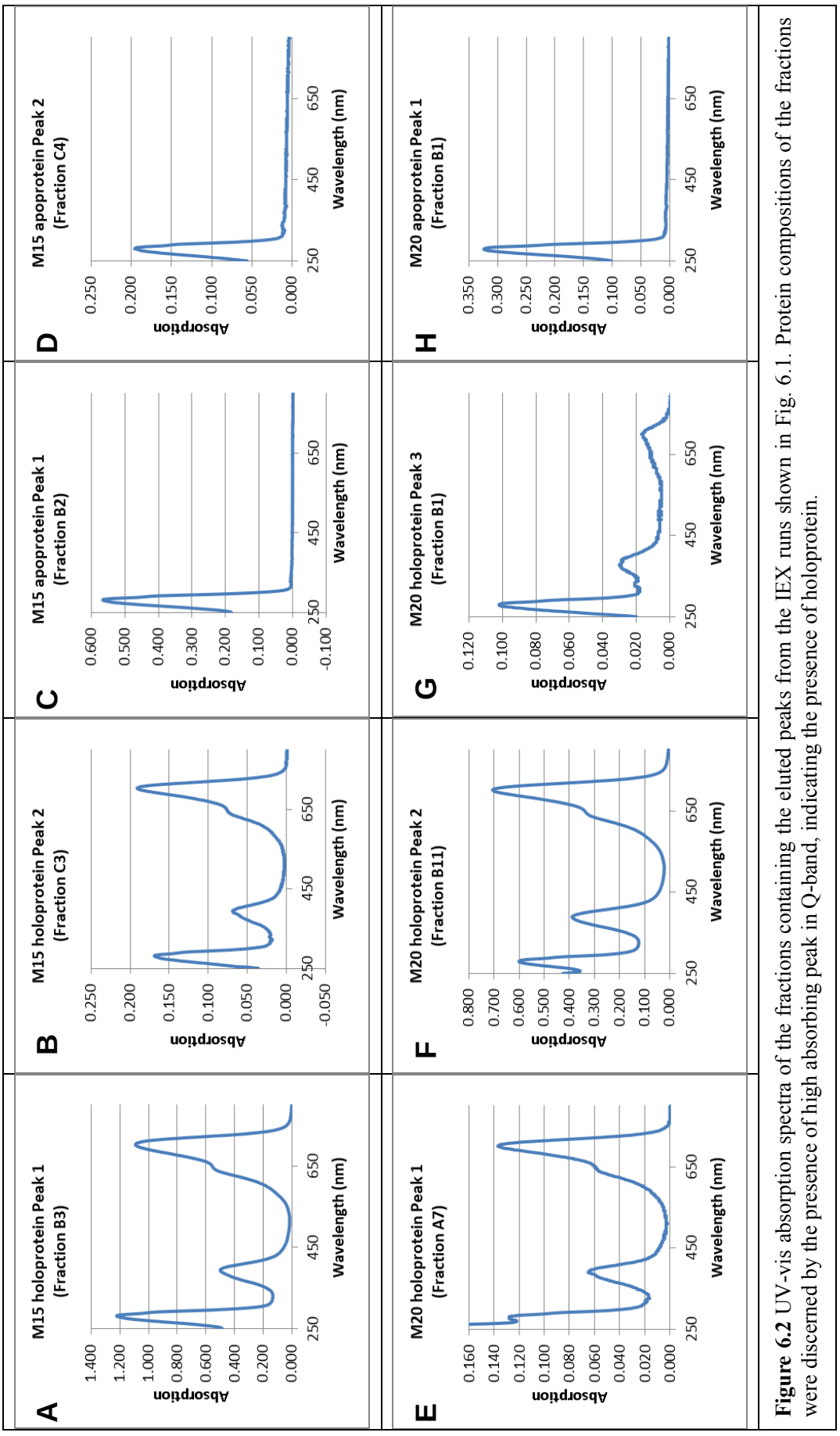


Figure 6.2 UV-vis absorption spectra of the fractions containing the eluted peaks from the IEX runs shown in Fig. 6.1. Protein compositions of the fractions were discerned by the presence of high absorbing peak in Q-band, indicating the presence of holoprotein.

6.2.2 M15s as apoprotein, BV adduct, and 15Ea adduct

IEX was also employed to probe the conformational change associated with photoconversion. To emulate the approach described in Section 6.2.1, Agp1 M15s fragments as apoprotein, BV-adduct or 15Ea adduct were subjected to analytical IEX.

3000 mAU×ml of Agp1 M15s as either apoprotein, BV adduct, or 15Ea adduct were applied and subjected to analytical linear gradient elution as described in Section 2.2.3.7.4. For the M15s apoprotein and the BV-adduct, qualitatively similar elution profiles to the previously observed ones (Section 6.2.1) were obtained (Fig. 6.3). Elution profiles of these proteins are characterised by the earlier main peak and the later minor peak that correspond to the monomer and the dimer, respectively. Both peaks of M15s-BV have smaller elution volumes than the corresponding peaks for M15s apoprotein. M15s apoprotein has a higher proportion of the later minor peak compared to M15s-BV. The elution profile of M15s-15Ea was characterised by a single peak with an elution volume of 58.73 ml which was presumed to correspond to the monomer peak of M15 apoprotein and the BV-adduct. In terms of the elution volumes, the difference between the 15Ea adduct and the apoprotein is smaller than the difference between the BV adduct and the apoprotein. Nonetheless, despite its low magnitude the difference in the elution volumes provides an opportunity to separate the 15Ea adduct from the apoprotein as demonstrated in Section 4.4.1. Overall, the differences in elution volumes of the main peaks amongst apoprotein, BV adduct, and 15Ea adduct are not drastic.

The fact that the 15Ea adduct elutes as a single peak whereas the BV adduct and the apoprotein elute as in monomer and dimer may enable to rationalise the relative kinase activity in terms of the proportion of the dimer in the sample. It has been reported that the relative kinase activities of Agp1 as apoprotein, in the Pr-form, and in the Pfr-form are 130%, 100%, and 19%, respectively (Inomata et al., 2005). Agp1 is a histidine kinase and formation of a parallel dimer is essential for its kinase activity. Determination of the area of peaks in the chromatograms in Fig. 6.3 indicates that the dimer peaks of the apoprotein, BV-adduct, and 15Ea amounts to 28.21%, 13.67%, and 0% of the total peak area, respectively. Interestingly, the decreasing order of the dimer peak proportions of these proteins accords with their relative kinase activities. Any model which describes the kinase activity derived from IEX data must be approached carefully, however. Firstly, these experiments were carried out with a truncated construct lacking the HK domain, and secondly the buffers of the protein samples were

adjusted to that of "IEX buffer A" with no salt prior to application to the IEX column. A buffer condition of low ionic strength with little physiological relevance may affect hydrophobic and/or ionic interactions at the dimer interface, therefore the proportion of dimer at the start of the analytical IEX may not reflect that in the *in vivo* condition. Therefore, while this idea is interesting, another technique must be employed to quantify the proportion of the dimer at biologically relevant ionic strength in order to make this hypothesis a valid one.

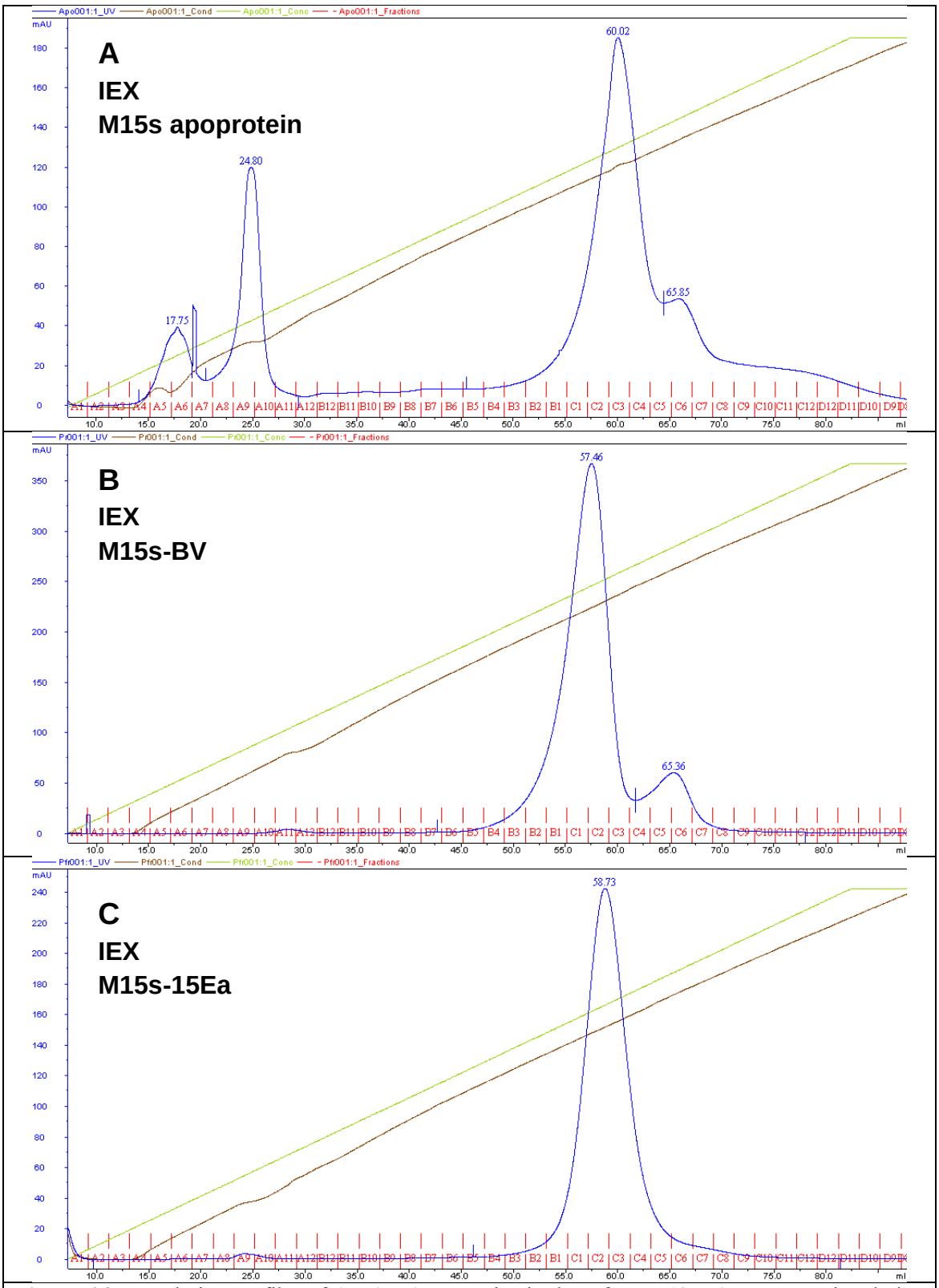


Figure 6.3 IEX elution profiles of Agp1 M15s proteins in three forms. A) M15s apoprotein, elution volume: 60.02 ml. B) M15s-BV holoprotein, elution volume: 57.46 ml. C) M15s-15Ea holoprotein, elution volume:58.73 ml. The blue, brown, and green lines indicate the absorption (mAU at 280 nm), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively.

6.2.3 Full-length protein as apoprotein, BV adduct and 15Ea adduct

Studies on conformational differences between forms of Agp1 with analytical IEX were extended to full-length Agp1. Full-length Agp1 includes a C-terminal HK domain which changes its activity according to the photochromic state of the phytochrome (Inomata 2005), therefore any conformational change in the HK domain may be detectable by IEX, especially since there was no major difference in the elution volume of the main peaks of M15 apoprotein, BV adduct, and 15Ea adduct.

1500 mAU×ml of full-length Agp1 either as apoprotein, BV adduct, or 15Ea adduct were eluted by a linear gradient from a HiTrap Q column (Section 2.2.3.7.4). Composition of the peak-containing fractions were analysed by absorption spectroscopy and SDS-PAGE (Fig. 6.5). The apoprotein and the BV adduct eluted at similar volumes of 60.53 ml and 60.09 ml, respectively, whereas the elution volume of the 15Ea adduct was distinctly different (47.11 ml; Fig. 6.4).

Given that no major differences in elution volumes were observed for three forms of M15s, the distinctly earlier elution of the 15Ea-adduct of full-length Agp1 is attributed to a unique property of the HK domain in the Pfr-form. The smaller elution volume of the 15Ea adduct may be caused either by a unique conformational change of the HK domain in the Pfr-form, altered oligomerisation state, or combination of both. In either cases the unique tertiary and/or quaternary structure of full-length Agp1 in the Pfr-form may help to rationalise its low kinase activity (19%) relative to the apoprotein (130%) and the Pr-from (100%) (Inomata et al., 2005).

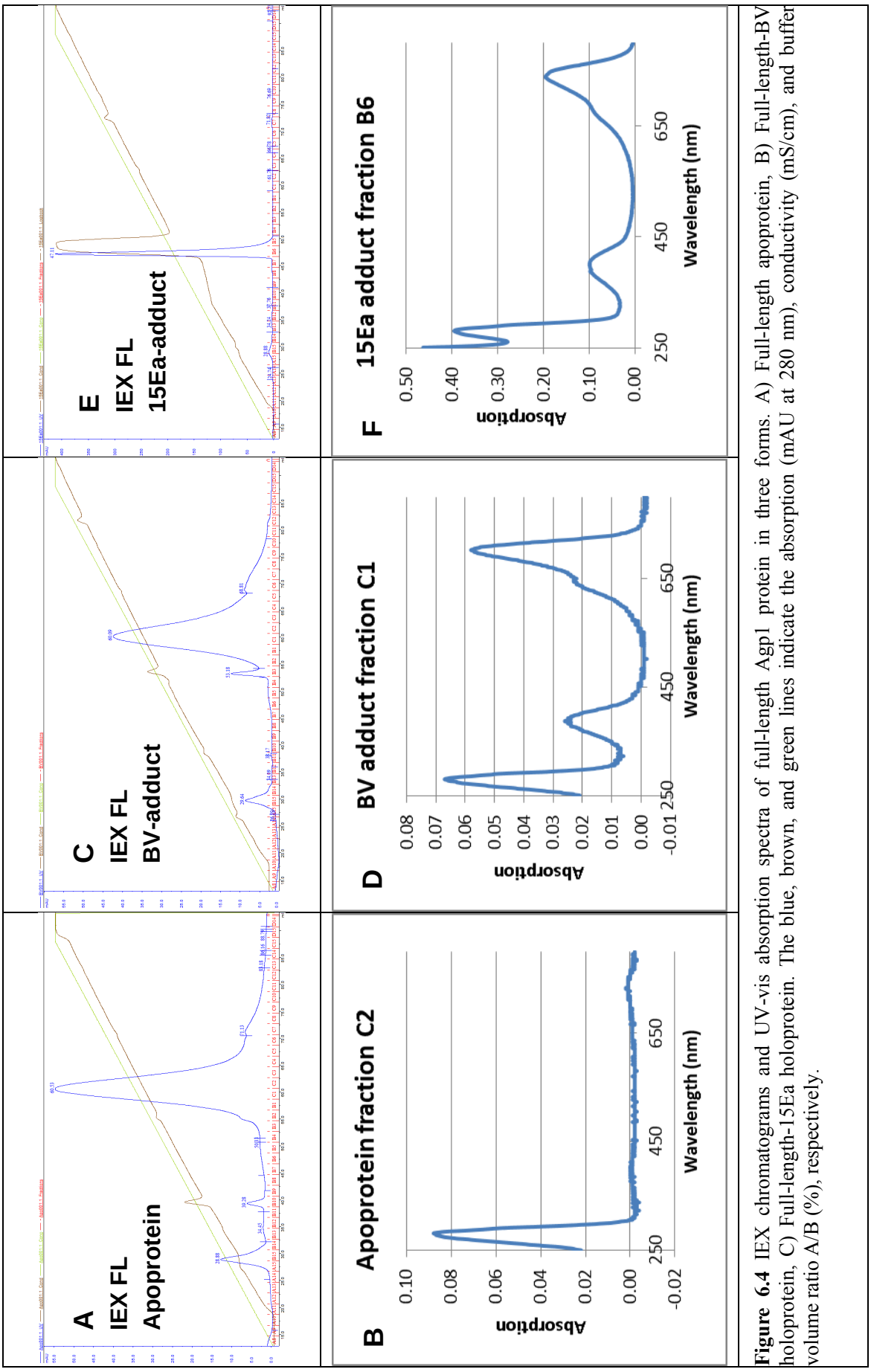


Figure 6.4 IEX chromatograms and UV-vis absorption spectra of full-length Agp1 protein in three forms. A) Full-length apoprotein, B) Full-length-BV holoprotein, C) Full-length-15Ea holoprotein. The blue, brown, and green lines indicate the absorption (mAU at 280 nm), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively.

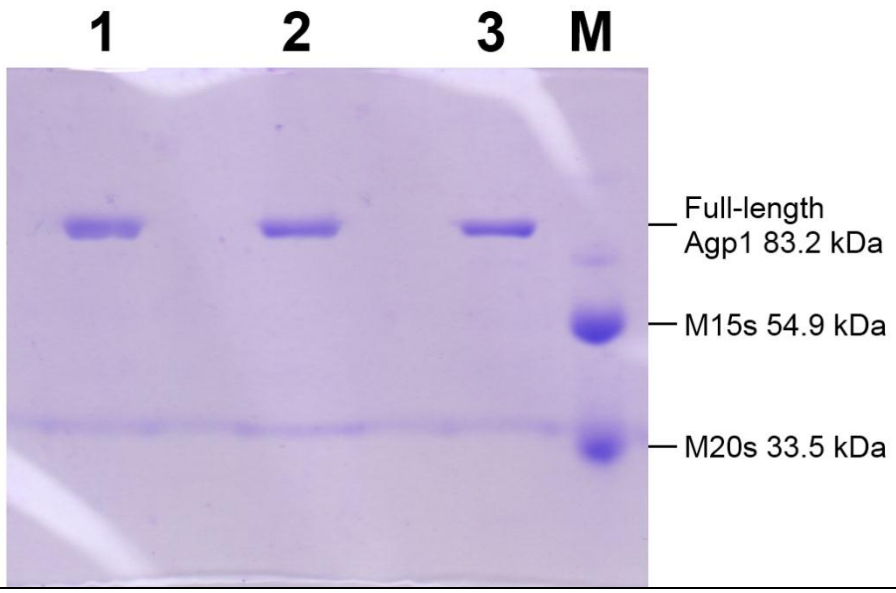


Figure 6.5 SDS-PAGE analyses of the fractions containing the main peaks of the IEX on full-length Agp1 proteins in three forms. Lane 1, Apoprotein fraction A4; Lane 2, BV adduct fraction A4; Lane 3, 15Ea adduct fraction B6; M, SN2 marker.

6.3 Native-PAGE

In addition to IEX, a complementary technique was deemed informative to probe the oligomerisation properties of Agp1. A rudimentary analysis of the oligomerisation properties of Agp1 was carried out by subjecting fragments of Agp1 in varying lengths to native-PAGE (Fig. 6.6). M15 and M20 proteins were tested as apoprotein and BV adduct. The first observation is that oligomerisation is observed in all combinations, although the extent of orders of oligomerisation depends more on the construct type (M15 and M20) rather than the form of the protein (apoprotein and BV-adduct). An oligomeric state of a higher order than the dimer is observed for M15, and oligomeric states of one order higher (trimer?) and two orders higher (tetramer?) can be clearly identified for the BV-adduct and the apoprotein, respectively. An oligomeric state of a higher order than the dimer is not clearly visible for M20 proteins.

Given the same construct and the same oligomeric state, the BV-adducts migrate faster for both M15 and M20 proteins under the same experimental condition (pH 8.8). This observation may be a contrast to the result of analytical IEX which suggested the net negative charges of apoproteins are greater than the BV-adduct in IEX buffers (pH 8.0) (Section 6.2.1). This apparent contradiction may be rationalised that in a native-PAGE gel, the rate of migration of proteins is affected by the protein conformation in addition to the net surface charge.

The conclusion that can be drawn from the analysis with native-PAGE is that the conformational change associated with BV assembly and its effect on the mobility in a native-PAGE gel is conserved between M15 and M20, therefore the likely site of conformational change is common between them, within the CBD (PAS-GAF domains).

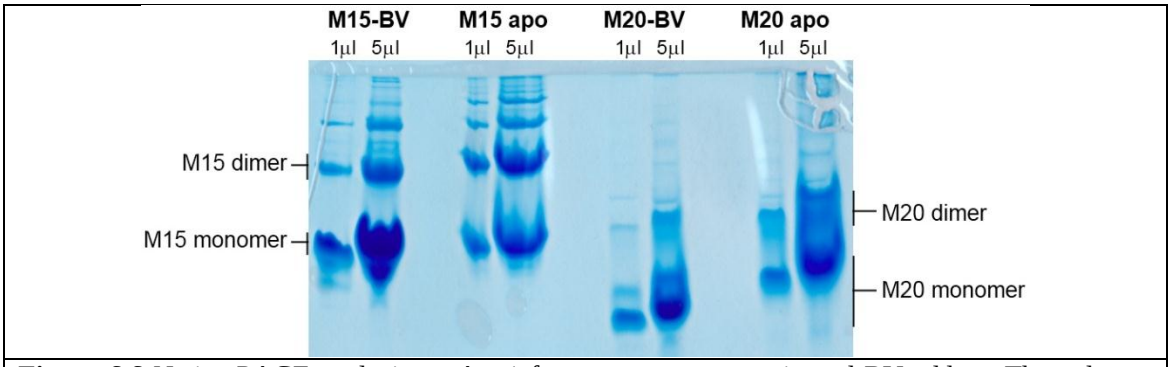


Figure 6.6 Native-PAGE analysis on Agp1 fragments as apoprotein and BV adduct. The volumes correspond to the volumes of the samples (5 mg/ml) per lane. The volumes of sucrose loading solution per lane were constant at 2 µl.

6.4 Hydrophobic interaction chromatography

HIC, like IEX, is a technique where proteins are separated according to the affinity of the proteins to the stationary phase of the column. The affinity in case of HIC is determined by the hydrophobicity of the protein surface. The hydrophobic interaction increases with higher ionic strength, and the protein is eluted with a buffer of lower ionic strength. A conformational change may involve a change in net surface hydrophobicity as well as in net surface charge, therefore HIC was also anticipated to provide information regarding the nature of conformational differences of Agp1 in different forms. HIC was employed to purify the DrBphP holoprotein from an inhomogeneous mixture, where the separation of proteins was presumably based on the same principle as discussed above (Auldridge et al., 2012).

6.4.1 M15s as apoprotein, BV adduct, and 15Ea adduct

M15s fragments of Agp1 as apoprotein, BV adduct, or 15Ea adduct were eluted by a linear gradient from HIC Buffer A to HIC Buffer B using a HiTrap Phenyl HP column (Section 2.2.3.7.5). The elution profiles of these three forms of the M15s protein were essentially identical, with all of them eluting at 100% HIC buffer B (20 mM Tris-HCl pH 7.8 at 20°C, 1 mM EDTA) (Fig. 6.7). The elution profiles indicate that the net surface hydrophobicities are equally high for all forms of the M15s protein. Similar elution volumes of the three forms of the M15s protein indicate that hydrophobicities of the protein surfaces of those proteins are similar despite conformational differences between them.

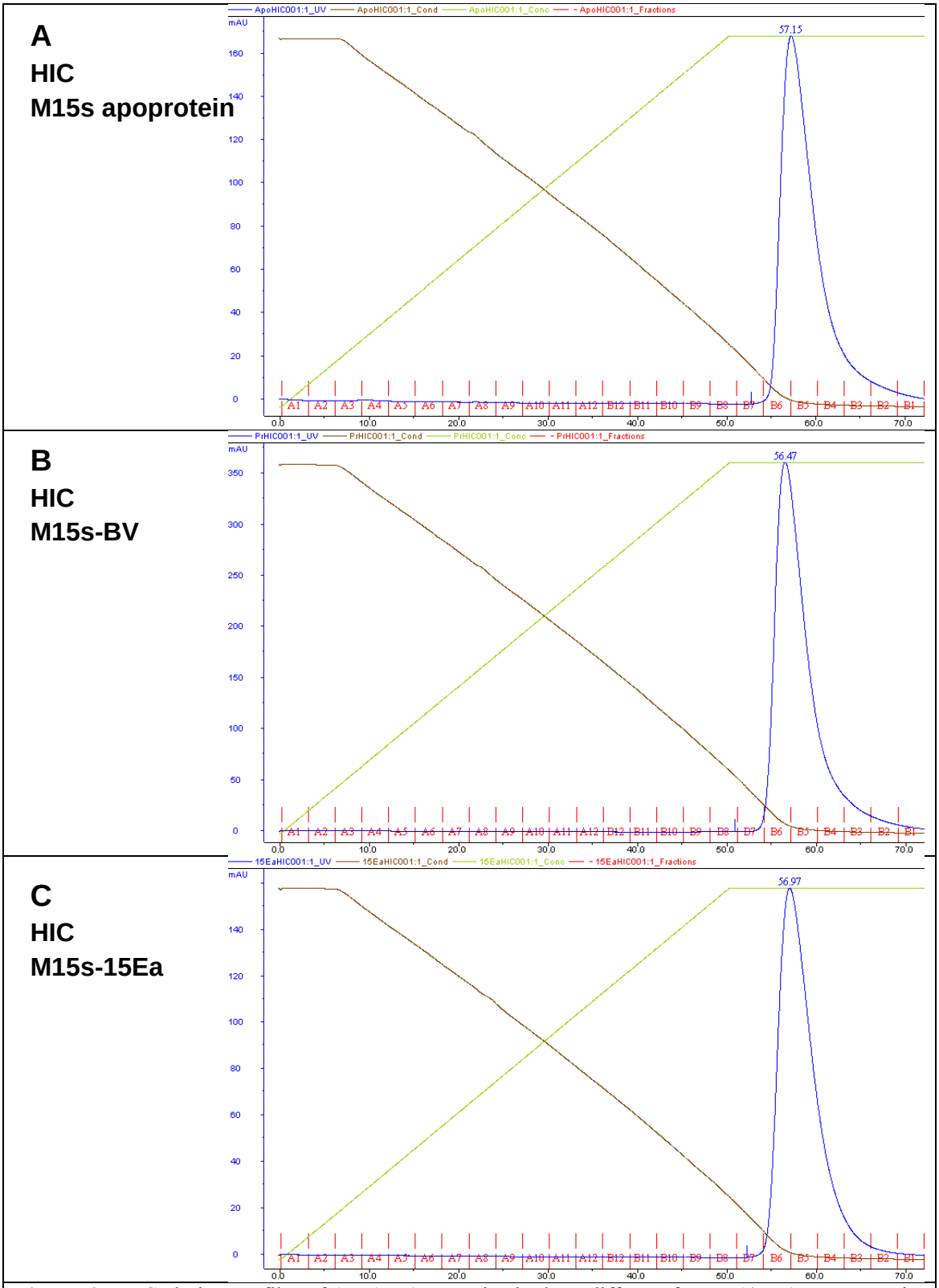


Figure 6.7 HIC elution profiles of Agp1 M15 proteins in three different forms. A) M15s apoprotein, 57.15 ml, B) M15s-BV holoprotein, 56.47 ml C) M15s-15Ea holoprotein, 56.97 ml. The blue, brown, and green lines indicate the absorption (mAU at 280 nm), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively.

6.4.2 Full-length protein as apoprotein, BV adduct and 15Ea adduct

In IEX, similar elution volumes were observed for the main peaks of M15s as apoprotein, BV adduct and 15Ea adduct, but substantial differences were found when the experiment was repeated with the full-length Agp1 protein which was interpreted to reflect the conformational difference of the HK domain. In order to investigate if differences in net surface hydrophobicity exist between apoprotein, BV adduct and 15Ea adduct of full-length Agp1, analytical HIC was also applied to full-length Agp1 proteins in these three forms using a HiTrap Phenyl HP column(Section 2.2.3.7.).

Similarly to the results from analytical HIC with M15s proteins, the three forms of full-length proteins all eluted at 100% HIC Buffer B, thus indicating the absence of substantial difference in the net surface hydrophobicity (Fig. 6.8). Analytical HIC could not reveal a conformational difference amongst them in a way that IEX could, and no further effort was made in the present study to utilise HIC to probe the conformational difference of the proteins.

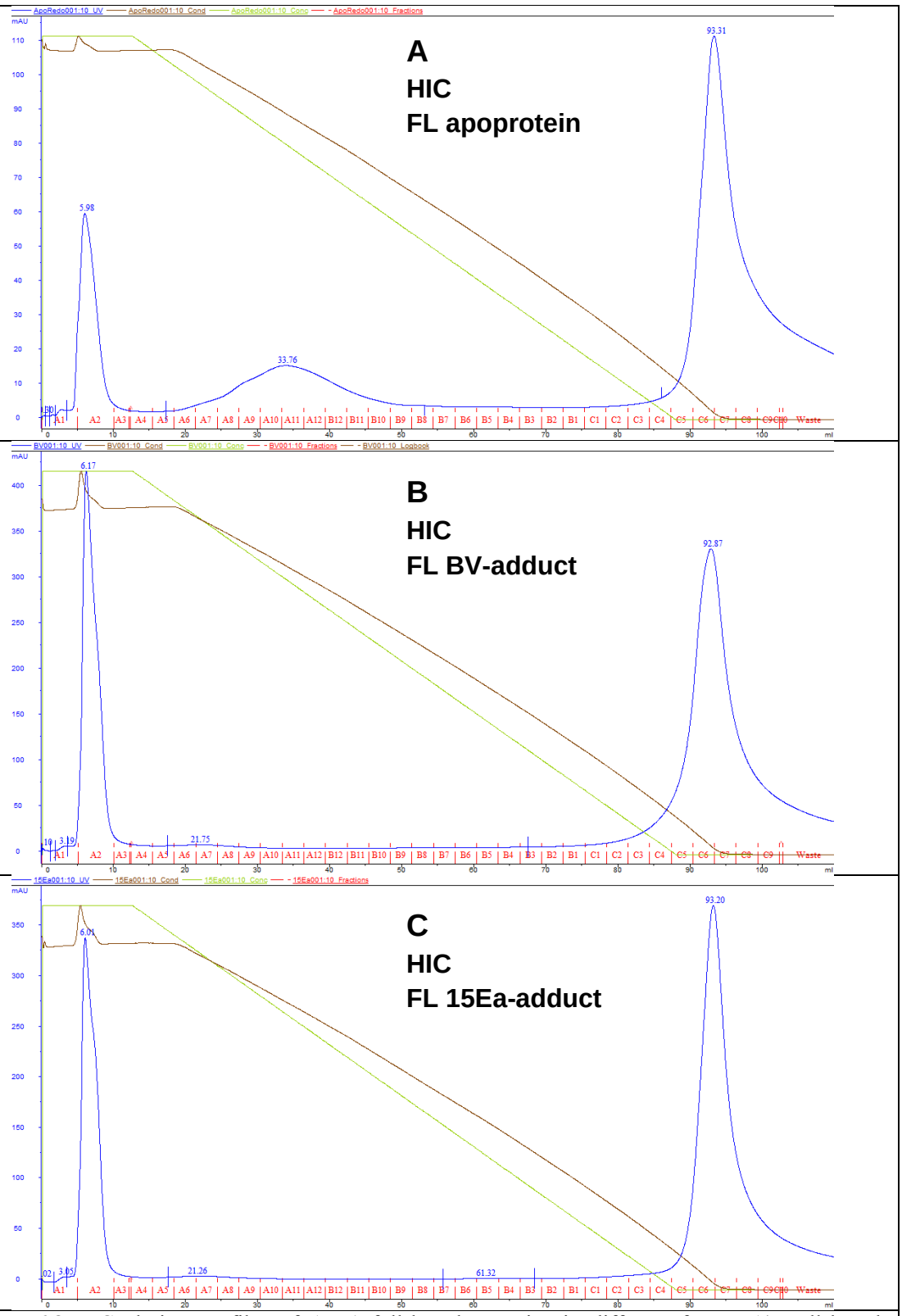


Figure 6.8 HIC elution profiles of Agp1 full-length proteins in different forms. A) Full-length apoprotein, B) Full-length-BV holoprotein, C) Full-length-15Ea holoprotein. The blue, brown, and green lines indicate the absorption (mAU at 280 nm), conductivity (mS/cm), and buffer volume ratio A/B (%), respectively.

6.5 Flexibility of the N-terminal region

The N-terminal region was assumed to be involved in a conformational change during the bilin assembly process because it shields the bilin from the solvent in the holoprotein structure. The N-terminal region also has an additional role of supplying the conserved Cys residue for the formation of a covalent bond with BV in bacteriophytochromes. Analysis of the N-terminal region of Agp1 may reveal the difference in functional roles played by this region in bacterial and cyanobacterial/plant phytochromes.

6.5.1 Apoprotein dimerisation via the N-terminal Cys residue

During a quality check of purified proteins with non-reducing SDS-PAGE, it was found that M15 and M20 apoproteins migrate unusually as dimers, but only if the free thiol was available at Cys 20 (Fig. 6.9A). Under the same condition, M15-BV and M15 C20A apoprotein migrate as monomers. The commonality between them is the lack of a free cysteine at the 20th position. A small amount of dimer was observed for M20-BV which had not been purified by IEX, therefore the presence of M20 apoprotein was likely. The M20 apoprotein consists of higher proportion of dimers than the M15 apoprotein. Interestingly, the persistence of the dimers under the presence of SDS indicates that a covalent bond mediates the dimerisation between the apoproteins. When the apoproteins that migrated as dimers under a non-reducing condition were subjected to reducing SDS-PAGE, both M15 and M20 apoproteins migrated as monomers, indicating that a disulphide bond mediates the dimerisation of apoproteins (Fig. 6.9B), thus confirming the earlier assumption that the free thiol group at the 20th position is the prerequisite for the formation of the dimer.

It was concluded that Agp1 apoproteins can dimerise via a disulphide bond at the conserved cysteine residue in the N-terminal region. The apoprotein dimerisation shown in Fig. 6.9 probably occurred as a result of gradual oxidation of the thiol group by oxygen during storage. The higher proportion of dimer in M20 than in M15 preparations after storage over a comparable period may be explained by the mode of protein-protein interaction being dominated by a larger natural interface in M15 rather than via disulphide bond. Oxidation of the thiol group is slow, and to this end, the samples were incubated under the presence of an oxidising agent, Potassium

hexacyanoferrate (III) (potassium ferricyanide, $K_3[Fe(CN)_6]$) to accelerate the oxidation prior to investigating the effect of an oxidising agent on Agp1.

Agp1 M15(s), the M15 C20A mutant, and M20(s) either as apoprotein or BV adduct at a concentration of 1 mg/ml were incubated for 15 minutes at 20°C in darkness under the presence of either 3 mM potassium ferricyanide or DTT to oxidise or reduce thiol group, respectively. Proteins were analysed by non-reducing SDS-PAGE (Fig. 6.10). Those proteins incubated with DTT migrated as monomers, and a minor difference in migration rate between the M20s apoprotein and M20-BV is attributed to the lack of first nine residues by the N-terminus in the former. The M15s and M20s apoproteins incubated with potassium ferricyanide migrated as dimers, whereas the proteins which lack the free cysteine at the 20^{th} position migrated as monomers. The presence of a small amount of monomer in the oxidised M15 apoprotein is attributed to insufficient incubation time. It can be ruled out that an insufficient amount of oxidising agent was responsible for incomplete oxidation of the thiol groups, since potassium ferricyanide was present in high molar excess at 3 mM, compared to the M15s apoprotein at 0.0182 mM. The presence of some dimer in the oxidised M20 holoprotein is attributed to a small contamination by some M20 apoprotein, as the sample used here had not been purified by IEX.

It is in general difficult to estimate the purity of the phytochrome holoprotein in a given sample by absorption spectroscopy. Since the formation of dimers via disulphide bonds at the conserved cysteine residue is specific to the apoproteins, detection of dimers may be utilised as a method to directly prove the presence of apoprotein in a sample. This method will be advantageous compared to using the SAR value as a criterion for purity which does not allow for an explicit quantification of the amount of the apoprotein in a sample. In order to test the applicability of this method to Agp1 proteins of various lengths, apoproteins of M15, M20, and full-length Agp1 at a concentration of 1 mg/ml were incubated with 4 mM of either DTT or potassium ferricyanide for 30 minutes before being subjected to non-reducing SDS-PAGE. Apoproteins of M20, M15, and full-length all formed dimers under the presence of the oxidising agent, thus demonstrating the applicability of this method using potassium ferricyanide in detecting apoproteins of various Agp1 fragments (Fig. 6.11).

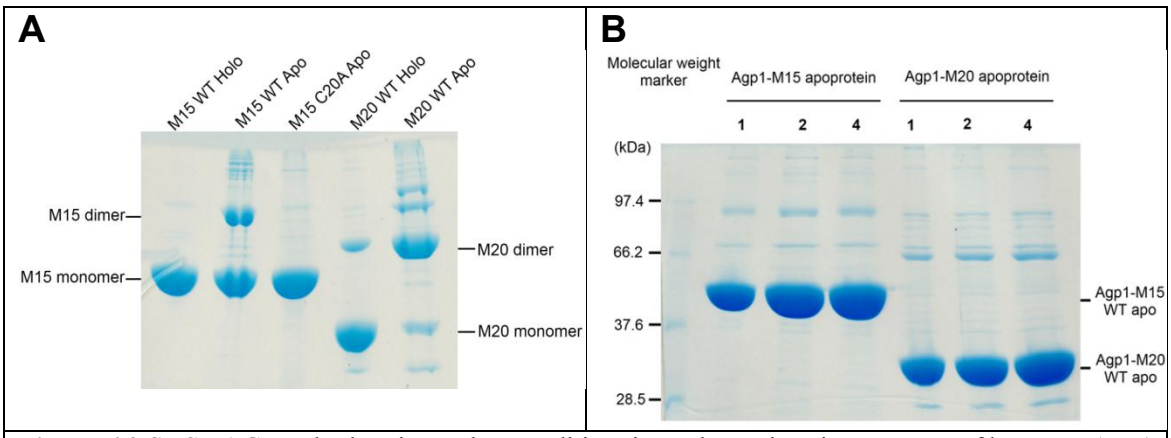


Figure 6.9 SDS-PAGE gels showing redox condition dependent migration patterns of by some Agp1 proteins. 20 μg of proteins were loaded per lane. A) Different forms of Agp1 proteins were subjected to SDS-PAGE with non-reducing SDS loading solution. Dimers of M15 and M20 apoproteins were detected despite the presence of SDS. Proteins with no free thiol group at Cys20 residue did not form dimer, except for M20-BV which contained an apoprotein contamination. B) Apoproteins of M15 and M20 which formed dimers were subjected to SDS-PAGE using reducing SDS loading solution in increasing quantity. The numbers indicate the volume of 4 × SDS loading solution in μl used per 20 μg of protein sample.

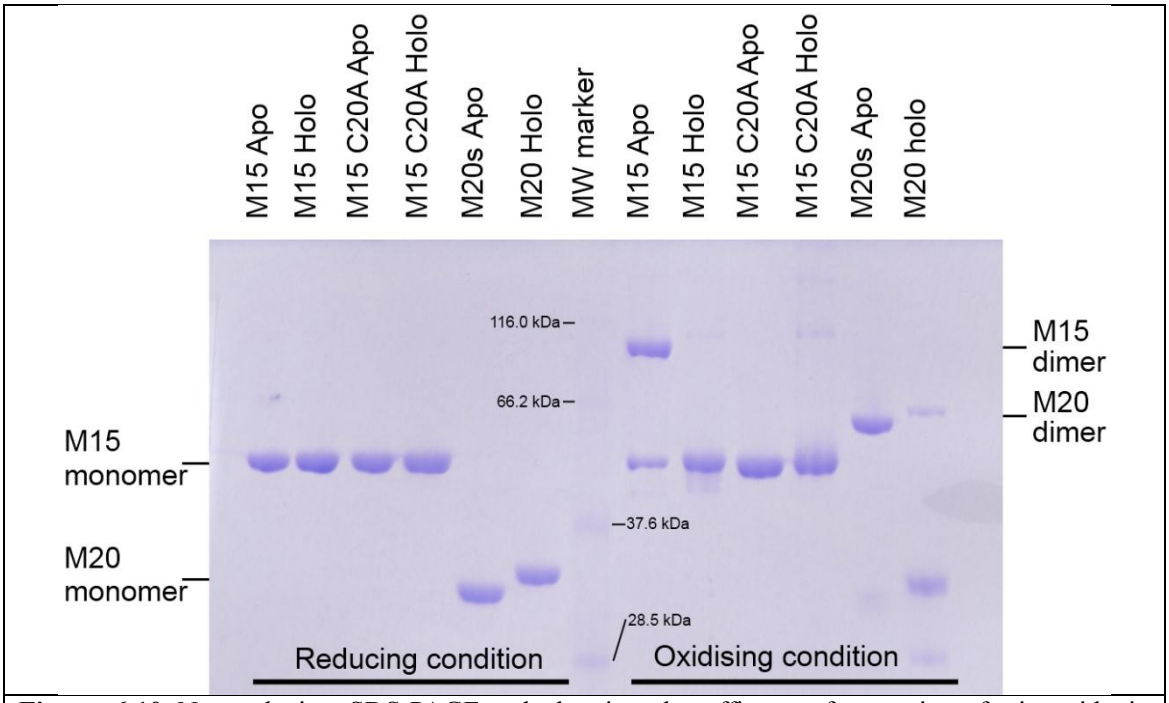


Figure 6.10 Non-reducing SDS-PAGE gel showing the efficacy of potassium ferricyanide in inducing the formation of apoprotein dimers via disulphide bond formation. 5 μg of protein was loaded with 2 μl of 4 × $SDS_{(non\text{-}reducing)}$ loading solution per lane. MW marker, Applichem marker 3 μl.

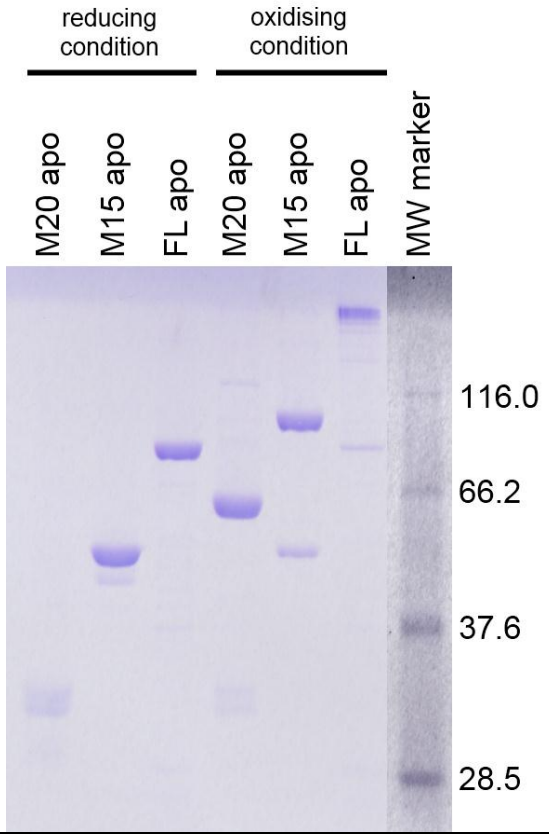


Figure 6.11 Non-reducing SDS-PAGE gel showing the occurrence of apoprotein dimerisation in 3 constructs of Agp1 of differing lengths. 5 µg of protein per lane. MW marker, Applichem marker 5 µl. The contrast of the MW marker lane was altered to aid visibility.

6.5.2 N-terminal region and bilin binding

To date, all phytochrome structures have been solved as the holoprotein. The holoprotein structures have in common that the GAF domain, and also the PAS domain if present, form the solvent shield round the bilin moiety, and it is obvious that the N-terminal region of the apoprotein must be in a conformation different from that of the holoprotein to allow for the initial bilin binding to occur. The inter-molecular disulphide bond at the N-terminal region (Section 6.5.1) is in accordance with the notion that this region in the apoprotein is flexible.

It is difficult to relate the conformational change of the N-terminal region and the mechanism of bilin assembly. The solvent exposed thiol side chain of Cys 20 may be hypothesised to recruit the bilin in the solvent by acting as the "bait", akin to proteins that function by the "fly-casting" mechanism (Huang and Liu, 2009). The available data contradicts such a model, however. It was shown in Section 4.2.2 that assembly of M15-15Ea involves the initial binding between the apoprotein and the bilin, followed by slow formation of the covalent bond, of which the latter leads to an increase in the absorbance in the Q-band and red-shifting of the absorption peak. Assembly with 15Ea is not a good representative of the process in nature, but a similar observation was made during the assembly of a non-covalent adduct, between the M15 C20A mutant and BV, which exhibited a blue-shifted peak and lower absorbance in the Q-band than the equivalent adduct with a covalent bond. Therefore, the assembly process is likely to consist of the initial binding followed by the formation of the covalent bond.

The role played by the N-terminal region has changed for some phytochromes during the course of evolution. In bacteriophytochrome, the N-terminal region shields the BV from the solvent as well as supplying the conserved Cys for the covalent bond. As the chromophore-binding Cys residue moved from the N-terminal region to the GAF domain in cyanobacterial and plant phytochromes, the N-terminal region was relieved from one of its functions. Sequence alignment of the N-terminal region reveals that the PG motif is unique to bacteriophytochromes (Fig. 6.12). This motif located beneath the "trefoil knot" (Wagner et al., 2005) may be functionally important for the binding between bacteriophytochromes and BV, but not for the binding of PΦB by plant phytochromes or PCB by cyanobacterial phytochromes. Structures of the region surrounding the PG motif in bacteriophytochromes and the equivalent in Cph1 reveal generally conserved molecular interactions despite dissimilar primary structure,

however. Interestingly, the PG motif is present in CEA001-RpBphP4, an achromo-phytochrome, although it does not even form a non-covalent adduct with BV. This could be because the presently functionally redundant PG motif in CEA001-RpBphP4 may not have been subjected to silent mutation yet.

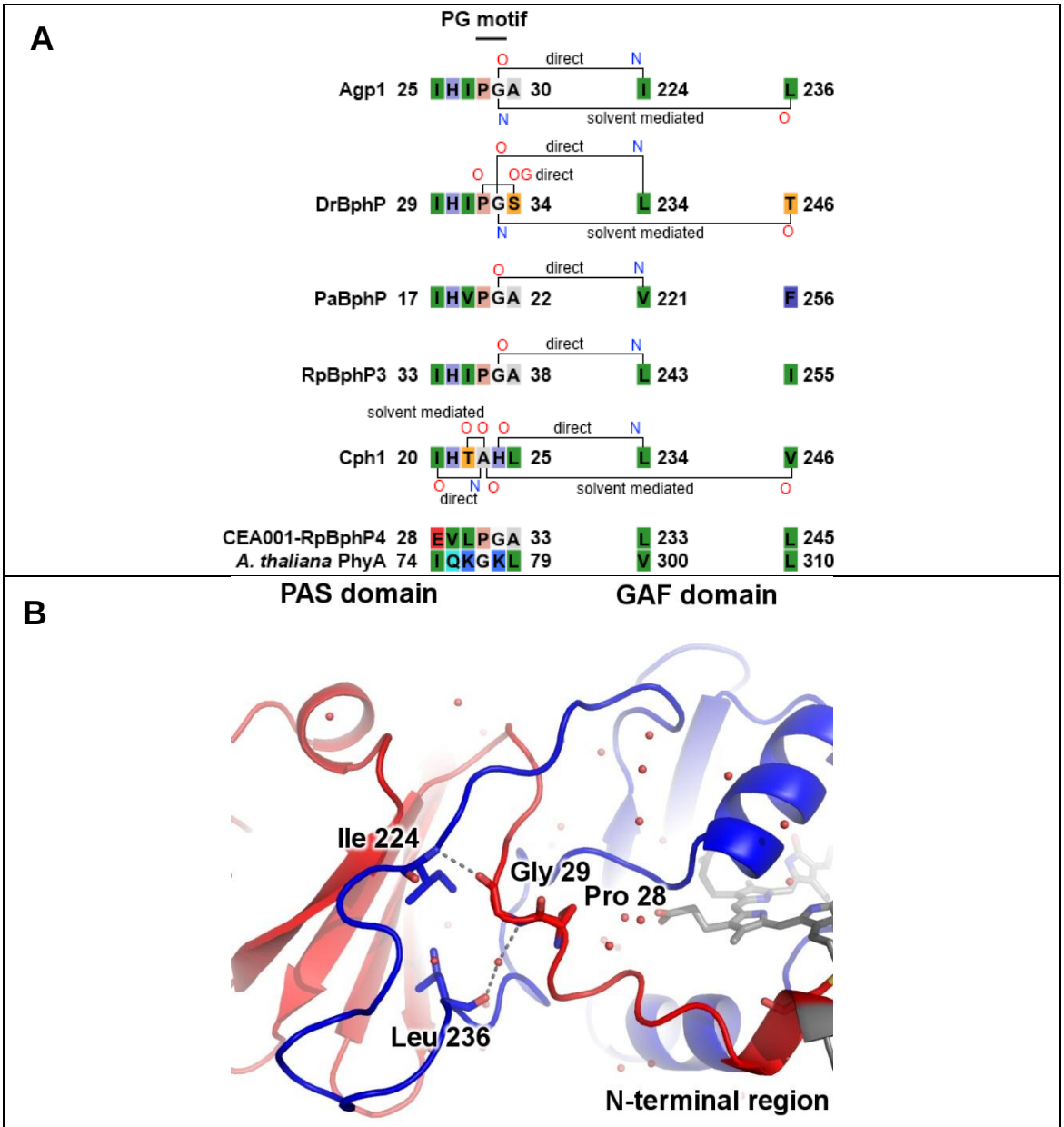


Figure 6.12 The PG motif in bacteriophytochromes. A) Sequence alignment of the PG motif and the interacting residues. In most cases, the molecular interactions are limited to main chain interactions and are mostly conserved between bacterial and cyanobacterial phytochromes. Functional importance of the PG motif in bacteriophytochrome is implied by its strong conservation, presumably because of its specific role in BV incorporation in bacterial, but not in plant and cyanobacterial phytochromes. The apparent lack of the equivalent interaction between Gly 29 and Leu 236 of Agp1 in PaBphP and RpBphP3 may be due to no mediating water molecule being modelled owing to their lower resolution data. Solid lines representing polar interactions are associated with a description of the type of interaction (direct or solvent mediated), and coloured alphabets indicate the atoms involved in the interaction. B) The PG motif in Agp1, shown as sticks along with the interacting residues. The protein is shown in cartoon representation. The PAS and GAF domains are coloured in red and blue, respectively.

6.6 Circular dichroism spectroscopy

A conformational difference between the apoprotein and the holoprotein of M15 could be detected by IEX, but IEX could not delimit the region of conformational difference (Section 6.2.1). To investigate if the conformational change during the bilin assembly process is associated with a change in the secondary structure content, CD spectra were obtained for M15s and M20s proteins as apoproteins and BV adducts. Spectral differences between the apoprotein and the holoprotein would indicate a change in the secondary structure during the assembly process. Furthermore, by comparing the degree of the difference between M15 and M20, the site of conformational change may be narrowed down because M20s is shorter due to the lack of the PHY domain. CD spectroscopy is advantageous over IEX, because it is not directly affected by the oligomeric state of the protein.

Also, based on the "context dependent conformation" of numerous conserved residues it was suggested in Chapter 3 that a dynamic conformational change is associated with the photoconversion from the Pr- to the Pfr-form. The hypothesis was tested by CD spectroscopic analysis of the M15s protein as BV and 15Ea adducts, representing the Pr- and the Pfr-forms, respectively.

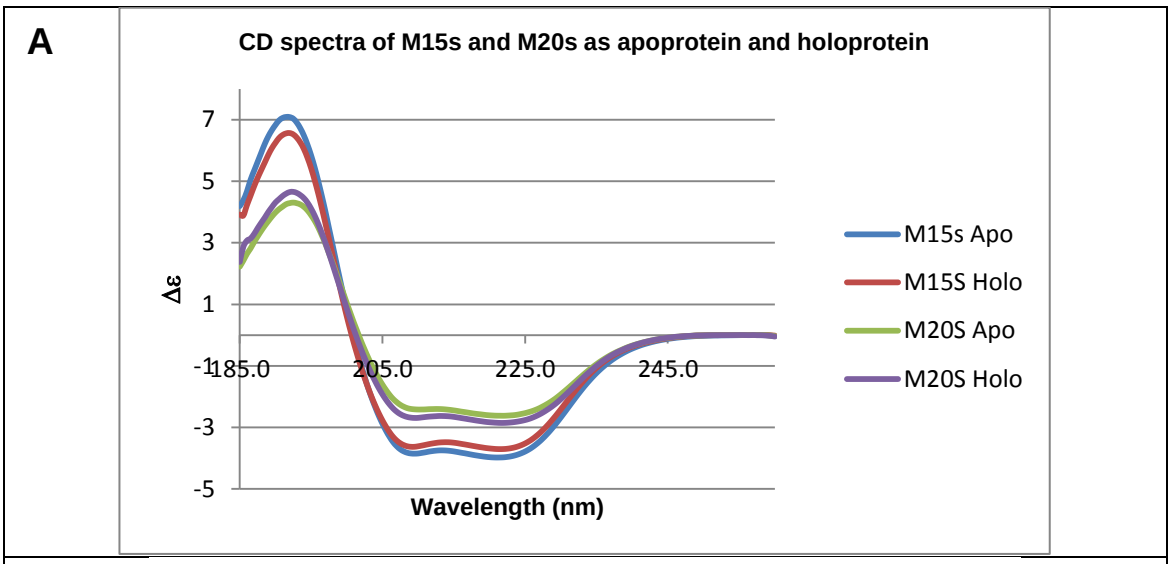
6.6.1 M15 and M20 as apoproteins and BV adducts

CD spectra of M15s and M20s fragments of Agp1 as apoproteins and BV adducts were obtained in the far-UV range (185 – 260 nm). Differences in the spectra were observed between apoprotein and BV adducts for both M15s and M20s, but the degree and the quality of the change was not consistent between M15s and M20s (Fig. 6.13A). This is possibly due to the errors introduced during the quantification of the protein concentrations.

Using the SER26-BV structure, the secondary structure contents of M15s-BV and M20s-BV were calculated, based on the assumption of structural isomorphism between the CBD and PCM fragments (Fig. 6.13C). A reasonable correlation was found between the secondary structure contents calculated from the CD spectra and the values calculated from the SER26-BV structure.

Overall, the differences in the secondary structure content between the apoprotein and the BV-adduct of the same fragment (M15s or M20s) are less significant than between

M15 and M20. Therefore, according to CD spectroscopic analysis the Agp1 apoprotein have a comparable degree of secondary structure elements as the BV-adduct.



B

Protein	α-helix	β-sheet	Turn	Disorder	Total
M15s Apo	0.38	0.16	0.18	0.28	1.00
M15s BV	0.37	0.18	0.19	0.27	1.01
M20s Apo	0.27	0.23	0.20	0.29	0.99
M20s BV	0.30	0.21	0.20	0.29	1.00

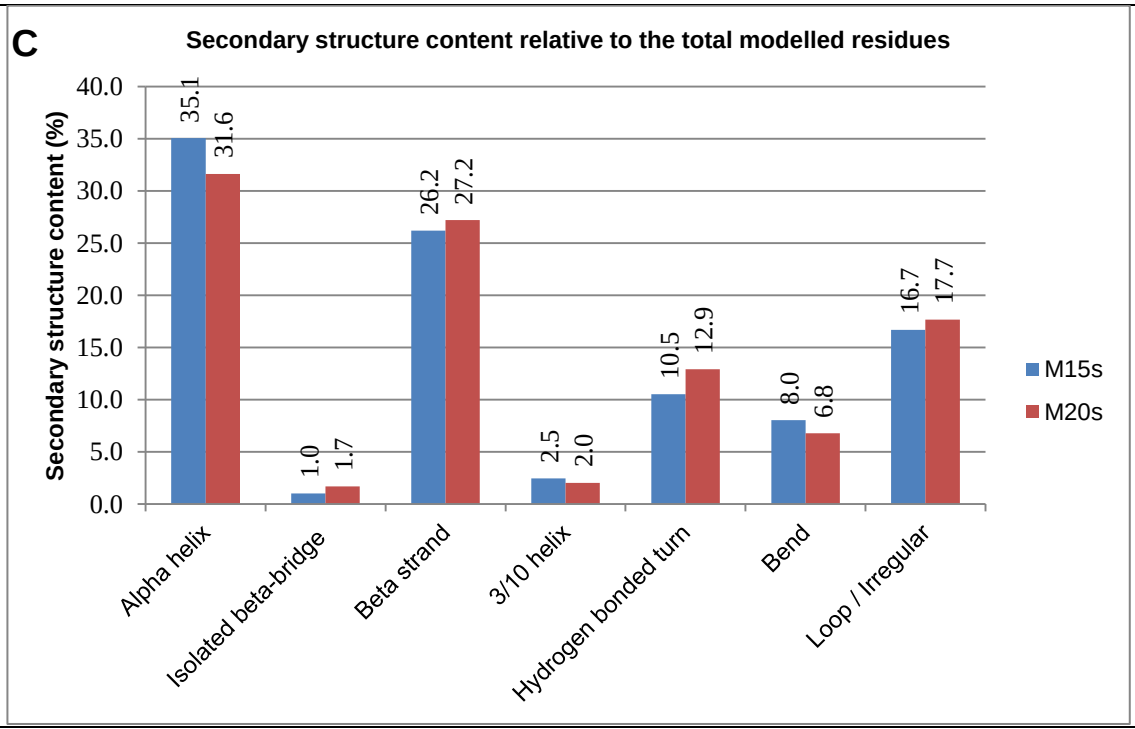


Figure 6.13 Secondary structure analyses on Agp1 M15s and M20s proteins. A) CD spectra of M15s and M20s proteins as apoproteins and BV adducts. Molar circular dichroism (Δε) is plotted as a function of wavelength (nm). B) Secondary structure contents of these proteins calculated using the DICHROWEB server. C) Proportion of secondary structure elements of the structure of SER26-BV, and of the M20 equivalent part of the structure. The calculation was carried out with DSSP.

6.6.2 M15 as BV and 15Ea adducts

In the crystal structure of phytochromes, conformational differences are found between those in the Pr- and the Pfr-form, for example the *Tongue* of the PHY domain comprises two anti-parallel β-sheets in the Pr-form and coils and an α-helix(ces) in the Pfr-form. Such differences manifest in the proportion of secondary structure elements in each structure (Fig. 6.14). The PaBphP structure in the Pfr-form features a lower proportion of β-strands and a higher proportion of loops than those structures which are in the Pr-form. The secondary structure content calculated from the structures, however, must only be regarded as a rough indication of the true proportion of the secondary structure elements. Crystallographic structures, especially the ones that were solved from lower resolution datasets often have un-modelled segments. The figures quoted in Fig. 6.14 are always shown as percentage relative to the total number of modelled residues, so the accuracy depends on the degree of disorder of the crystal structures.

If the conformations found in the Pr- and the Pfr-forms of phytochromes represent the two end points of the photoconversion trajectory, such changes may be detected by CD spectroscopy, and to this end, CD spectra of M15s as BV adduct and 15Ea adduct representing the Pr- and the Pfr-form, respectively, were collected (Fig. 6.15).

Only slight changes in proportions of secondary structure element were found between the BV and 15Ea adducts of the M15s protein. The 15Ea adduct representing the Pfr-form showed a 1% higher content of α-helix, a 1% lower content of disordered regions, and the same proportion of β-sheet structure when compared to the BV adduct. These data are difficult to relate to the differences in the secondary structure contents of phytochrome crystal structures in the Pr- and the Pfr-forms. Changes in the secondary structure contents by few per cents in a relatively large protein (Agp1 M15s = 54 kDa) are close to the limitation of CD-spectroscopy, because the results are greatly affected by an inaccuracy during the quantification of the protein concentration.

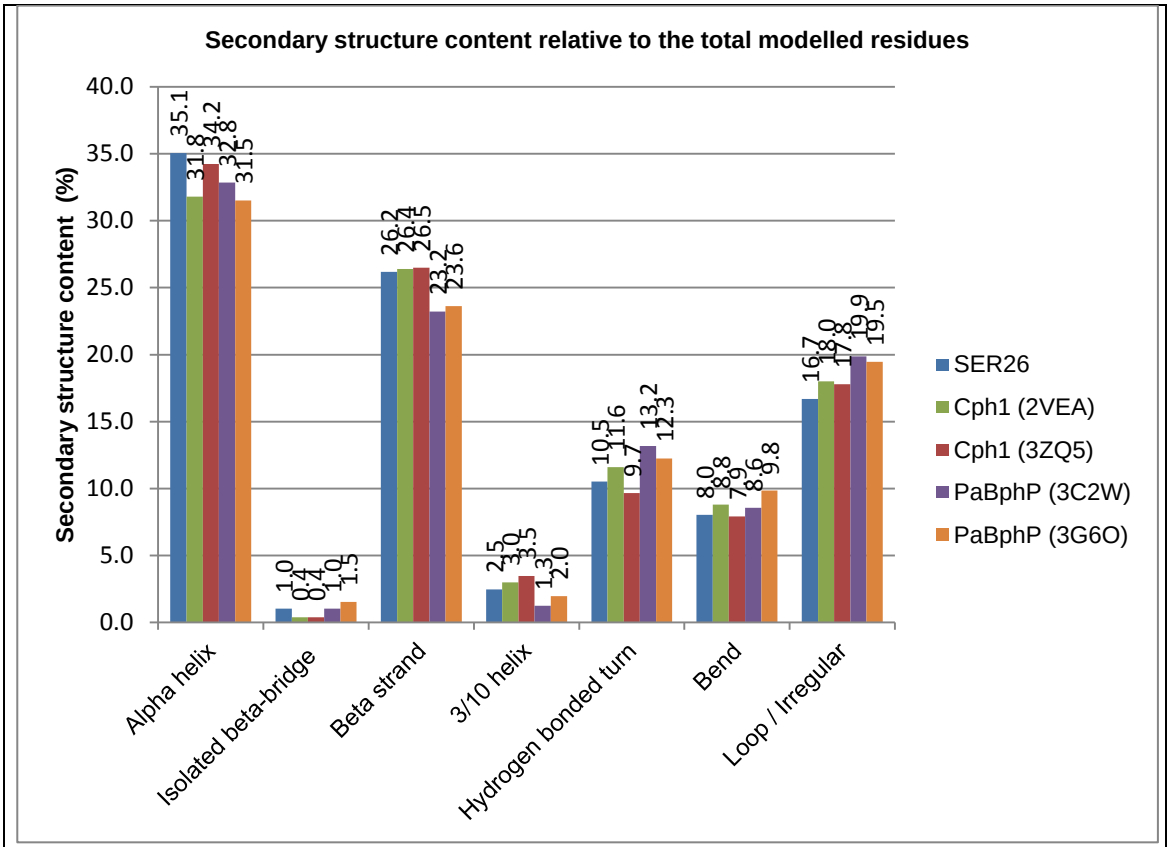
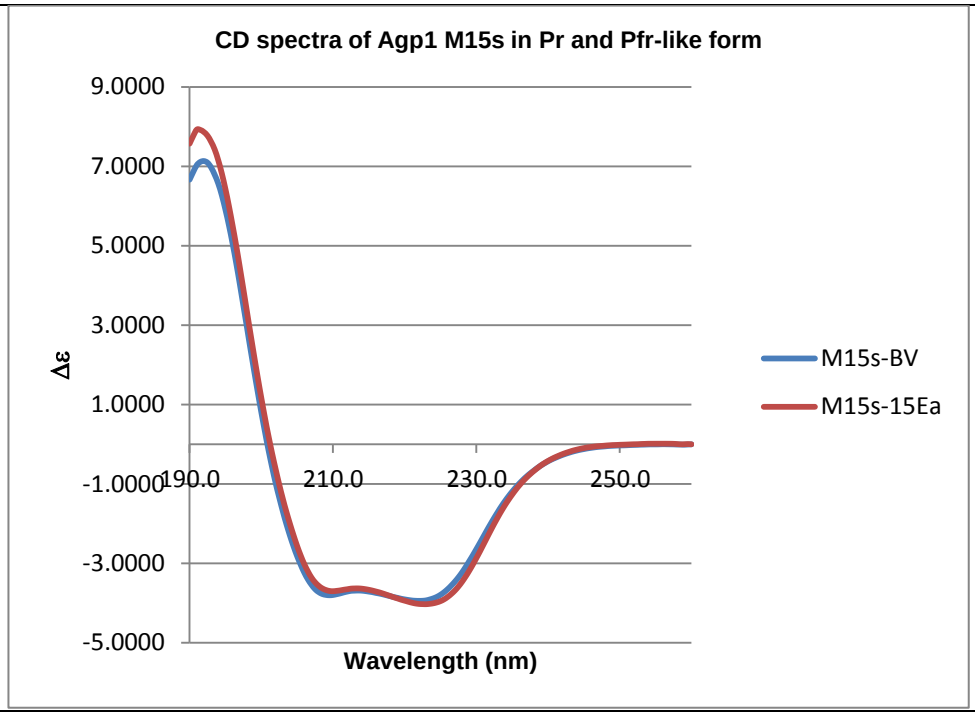


Figure 6.14 Proportion of secondary structure elements in phytochrome PCM structures. Calculations were performed with DSSP, and the proportion is relative to the total number of modelled residues. Care must be taken to interpret the secondary structure contents, because the degrees of disorder that affect the values differ between the structures. The structures used for the analysis and their PDB accession codes are as follows. Agp1 SER26-BV (N/A), PaBphP (3C2W and 3G6O), Cph1 (2VEA and 3ZQ5). The PaBphP structure with PDB code 3IBR was omitted due to the lower quality of the data and a very similar proportion of secondary structures as in the structure with PDB code 3G6O.

A



B

Protein	α-helix	β-sheet	Turn	Disorder	Total
M15s-BV	0.39	0.16	0.17	0.27	0.99
M15s-15Ea	0.40	0.16	0.18	0.26	1.00

Figure 6.15 Secondary structure analysis of the M15s protein in the Pr and the Pfr-like forms. A) CD spectra of M15s-BV and M15s-15Ea presented as molar circular dichroism (Δε) plotted as a function of wavelength (nm). Higher absolute values at λ ≈ 193 nm and λ ≈ 225 nm in the 15Ea adduct indicates a higher proportion of α-helices. B) Secondary structure contents of M15s-BV and M15s-15Ea calculated using DICHROWEB server.

6.7 Discussion and future works

6.7.1 Conformational differences between different forms of Agp1

Combinations of biochemical approaches have revealed some aspects of conformational differences between Agp1 in different forms. The flexibility of the N-terminal region is suggested to facilitate the initial diffusion of BV into the bilin pocket. M15 fragments as apoprotein, BV- and 15Ea adduct did not show major differences in IEX elution volumes, but when the experiment was repeated with the full-length protein, the elution volume of the 15Ea adduct was distinctly different. This larger elution volume by the 15Ea adduct is likely to be a conformational difference at the HK domain, and this in turn could explain the lower kinase activity by the Pfr-form compared to other two forms.

CD spectroscopy on M15 and M20 fragments of Agp1 as apoprotein and BV-adduct has shown that no substantial difference in proportions of secondary structure elements is present between BV-adduct and apoprotein. This indicates that phytochrome apoprotein structure, which to date little detail is known, features similar secondary structure contents to that of the BV adduct, both as M20s and M15s fragments. The apoprotein therefore, is as well-folded as the BV-adduct holoprotein, and conformational differences detected by IEX is due to changes in the tertiary structure rather than the secondary structure.

CD spectroscopy on M15 proteins as the BV- and the 15Ea-addcut did not indicate a major difference in the secondary structure content between them, at least not in the scale of the difference between phytochrome PCM crystal structures in the Pr- and Pfr-form. This is experimental evidence against a hypothesis that phytochrome crystal structures in the Pr- and the Pfr-forms represent the end points of the photoconversion (Section 3.11.4). Perhaps the conformational change during the photoconversion involves changes at the tertiary structure level and relatively small changes in the secondary structure. Observations such as the difference in SEC elution volumes of phytochromes in the Pr- and the Pfr-forms (Section 5.3.2) may be rationalised by a difference in the global conformational differences (Section 3.10.2), but without a structural evidence, all proposals remain to be a mere speculation. Therefore it is ever more important to solve the structure of phytochrome in its photoconverted state as well as its ground state.

6.7.2 Oligomerisation properties of Agp1

Analysis with native-PAGE indicated two aspects of the oligomerisation properties of Agp1.

The first point is that the affinity between monomers is such that in a native-PAGE, a larger proportion of protein was found in the monomeric form than in the dimeric form. This is consistent with the result from IEX runs using M15 fragments (Section 6.2.1). Seemingly weak affinity between monomers to form a dimer may be correlated with the fact that Agp1 lacks an Arg residue at the 131th position which is implied to play a key role in dimerization (Bellini and Papiz, 2012a).

Secondly, Agp1 M15 oligomers of clearly higher order than the dimer were observed. This is consistent with some previous observations (Noack et al., 2007, Scheerer et al., 2010). Since the order of oligomeric states increases sequentially (Dimer, trimer and then tetramer, Noack et al., 2007), the observation is unlikely to derive from non-specific inter-molecular aggregation. No functional role of the higher order oligomeric states of phytochrome has been characterised. Current lack of information regarding the molecular interface of higher order oligomers does not permit a study of the nature of the interface using an electrostatic surface potential of Agp1 (Figure 3.31). Since the experiment with a native-PAGE gel was carried out with the M15 fragments, the formation of higher order oligomers is mediated by the N-terminal region of Agp1. Higher order oligomers were observed in the M15 fragments in the present study, but neither the data presented here nor a study by Noack et al. (2007) exclude a possibly for the C-terminal HK to also mediate the formation of higher order oligomers in the full-length Agp1 protein.

6.7.3 Hydrophobic interaction chromatography

No major differences in elution volumes were found between the apoprotein, BV adduct, and 15Ea adduct of Agp1 with HIC, regardless of the length of the protein used (full-length or M15). Therefore it was not possible to purify the Agp1 holoprotein from the apoprotein with the HIC column, but this is in contrast from the reported case where BV adduct of DrBphP CBD was successfully purified with a HIC column with the same functional group (phenyl sepharose column, Auldridge et al., 2012). Difference between Agp1 and DrBphP may be attributed to the difference in the lengths of the proteins used during HIC purification, but it is also possible that the degree and the nature of the

change at the protein surface during bilin assembly to differ between phytochromes. This hypothesis may be confirmed by measuring the elution volume of Agp1 M20s protein as apoprotein and BV adduct.

6.7.4 Disulphide bond formation

Intermolecular disulphide bond formation was shown to form via Cys 20, but this mode of dimerisation was regarded to be physiologically irrelevant, at least in *A. tumefaciens*. Under the native condition in *A. tumefaciens*, the phytochromes are only expressed at low concentrations of 10 and 19 molecules per cell for Agp1 and Agp2, respectively (Oberpichler et al., 2006). Given the continuous production of BV by heam oxygenase, BV molecules are assumed to saturate the free thiol groups of apo-phytochromes.

CEA001-RpBphP4, an achromo-phytochrome that functions as a redox-sensor which does not even form a non-covalent adduct with a bilin has been reported to undergo redox-state dependent oligomerisation via Cys residues (Vuillet et al., 2007). The way CEA001-RpBphP4 apo-phytochrome forms oligomer via disulphide bond is superficially similar to the intermolecular disulphide observed in Agp1 apoproteins, but functional role of oligomerisation by these two phytochromes are expected to be different. For example, the positions corresponding to Cys 472 and Cys 722, the Cys residues that are implied to be important for redox-state dependent oligomerisation in CEA001-RpBphP4 are usually occupied by Ser and Leu, respectively, in chromophore-binding bacteriophytochromes including Agp1.

Chapter 7: Overall summary

In the present study, the structure of Agp1, a bacteriophytochrome, was solved as a photosensory core module fragment in the Pr-form. As far as is known, this is the first time that the entire photosensory core module fragment of a classical bacteriophytochrome has been structurally determined.

The structure of Agp1 resembles the cyanobacterial phytochrome Cph1 in terms of its global conformation. The similarity is significant considering the differences between those two phytochromes with regards to the type of the bound bilin chromophore and the location of the chromophore-binding cysteine residue, and as expressed by the distance in molecular phylogeny. Therefore we can now be confident that the global conformations of these structures are valid representatives of the Pr-form of phytochromes as a whole.

Molecular interactions found in Agp1, on the other hand, resemble the bacteriophytochrome DrBphP more closely than Cph1. This is thought to correlate with the functional differences between cyanobacterial phytochromes and bacteriophytochromes, such as the ability of bacteriophytochromes to undergo dark-reversion from the photoconverted state into the dark-adapted state, which is absent in cyanobacterial phytochromes.

An open question remains as to whether the bathy-phytochrome structures are suitable to represent key features of the photoconverted state of the classical phytochromes. There are now two representatives of bathy-phytochrome structures, and they too show structural similarities to each other despite a significant phylogenetic distance between them. Furthermore, some residues that stabilise their structures in the Pfr-form are not conserved. Structural similarities between two phylogenetically distant bathy-phytochromes reinforce the notion that their global conformations represent the Pfr-form for all phytochromes. The situation leaves open the possibility for an alternative hypothesis that the similarities between two bathy-phytochromes structures do not represent the photoconverted state of classical phytochromes and that the similarities are the result of convergent evolution.

A comparison of the structures of phytochromes in the Pr- and the Pfr-forms reveals many instances of “context dependent conformation”, a phenomenon where residues conserved in both classical and bathy-phytochromes exhibit different conformations in

two forms. This observation supports the notion that a phytochrome whose structure is known in one form can assume the conformation of the other, and supports the first of the two hypotheses presented above – that the crystal structures that are currently available for the Pr- and the Pfr-forms truly represent the end points of the photoconversion event for all phytochromes in terms of their global conformation.

Appendix

Appendix 1: The safe-light

The safe-lights consisting either of filters or and LED have the peak in transmission spectra and emission spectra, respectively, at the wavelength close to an absorption minimum of Agp1 sample (Fig. A.1). This in theory should minimise photoconversion from the Pr-form into the Pfr-form, and that was also the case as far as observed during this study.

Appendix 1.1 Green filter

Colour effect filters, Velvet green #735 and Twickenham green #736 (LEE filters) were used either by themselves or in combination to cover light emitting materials such as power indicator of instruments or monitor display in order to minimise the exposure to light at photoactive wavelengths by phytochrome samples. Those filters were also used to cover handheld torches (Halfords) (Fig. A.1), and to cover the lamp of a stereo microscope during inspection of crystallisation plates.

Appendix 1.2 Green LED

NSPG510AS, a green LED with the maximum emission at 525 nm (Nichia) was connected in serial with a register rated at 30 ohm, and then 10 of such combinations were connected in parallel and were powered by an adjustable AC/DC converter set to operate at 3.5 volts. The resulting safe lamp is shown in Fig. A.1 and the circuit diagram is shown in Fig. A.2.

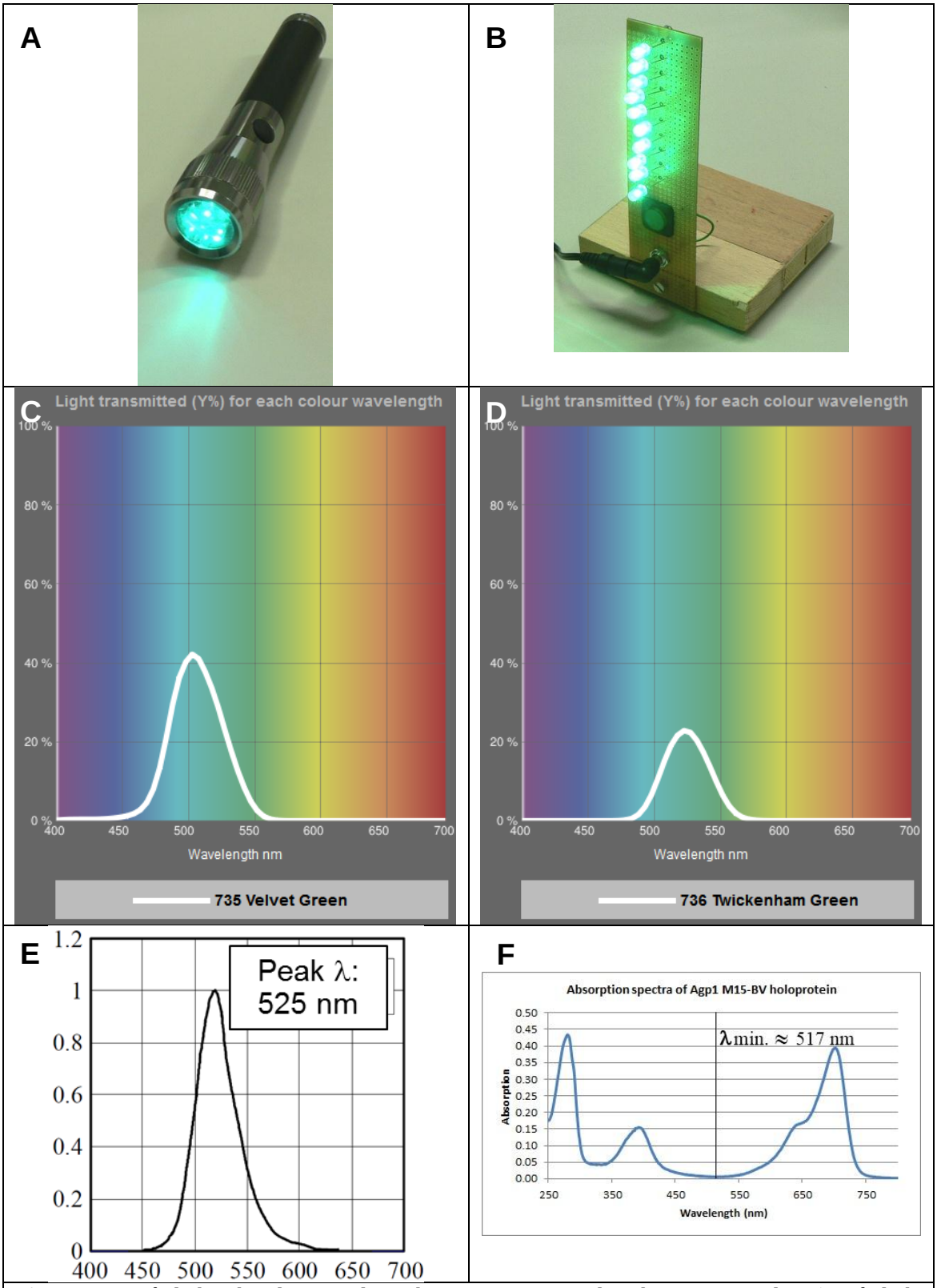


Figure A.1 Safe lights that do not induce photoconversion used in the present study. A) Safe light constructed with a green filter and a torch. B) Safe light constructed with green LEDs. Transmission spectra of C) #735 Velvet green filter and D) #736 Twickenham green filter (Lee Filters). E) The emission spectra of NSPG510AS green LED (Nichia). F) Typical absorption spectra of Agp1-BV in the Pr-form. An absorption minimum is indicated by the vertical line, which is roughly at 517 nm

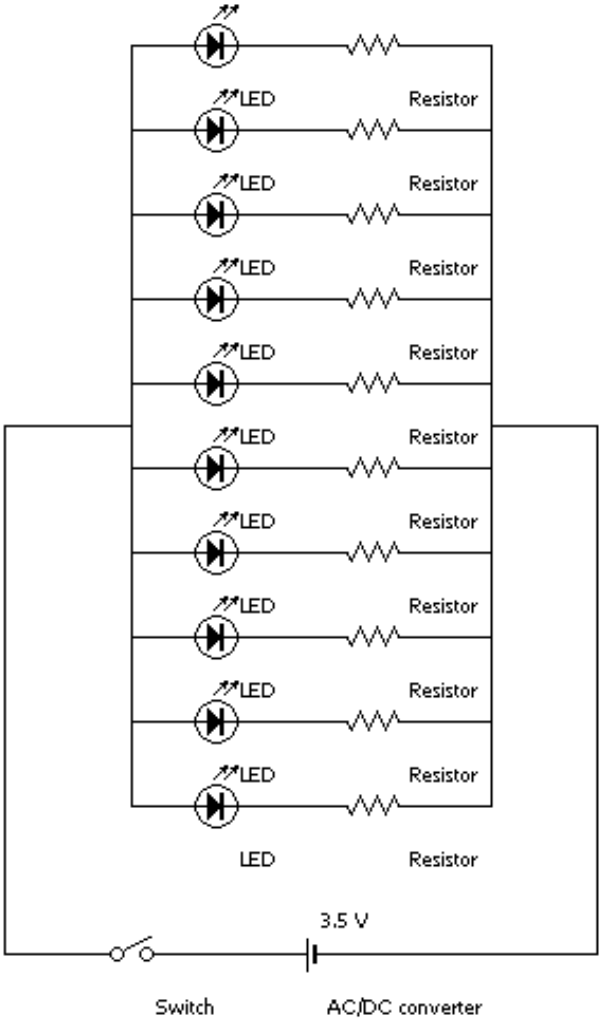


Figure A.2 The circuit diagram of the safe-light used in the present study. The photograph of the assembled equipment is shown in Fig. A.1. The emission spectrum of the green LED (Nichia NSPG510AS) has the peak at 525 nm which nearly coincides with a minimum of the absorption spectrum of Agp1-BV holoprotein (Fig. A.1). The resistors were rated at 30 Ω. The total current during operation at 3.5 V is 0.143 A, which is below the capacity of the AC/DC converter (0.5 A). The current though each LED is 0.0143 A, which is below the maximum forward current of the LED (0.035 A). The diagram was created with BSch3V (Suigyodo).

Appendix 2: Generation of M20s plasmid construct

Appendix 2.1 Polymerase chain reaction and T4 DNA ligation

PCR reaction was carried out with pNM19a plasmid bearing the PAS, GAF and PHY domains as the template. pNM19a-M20F and pNM19a-M20R primers were used (Section 2.2.2.1). PCR primers were designed to frank the sequence to be omitted (the PHY domain) and to retain the C-terminal hexahistidine tag to enable affinity purification of the expressed protein (Fig. A.3).

The amplified linear DNA fragment was purified with DNA purification kit and subjected to blunt end ligation with T4 DNA ligase at 20 °C for 1:30 hours. The ligation reaction was inactivated by incubation at 65°C for 10 minutes. The sample was purified with DNA purification kit again prior to being subjected to DpnI digestion. Only the parental pNM19a plasmids, which had been purified from *E. coli* cells, are expected to be digested by DpnI enzyme due to their methylated state. *E. coli* XL1-Blue cells were transformed with the DNA sample and subjected to selection with LB-agar plate containing ampicillin. Several colonies were selected for plasmid DNA purification with Miniprep kit, and subjected to verification for the successful truncation.

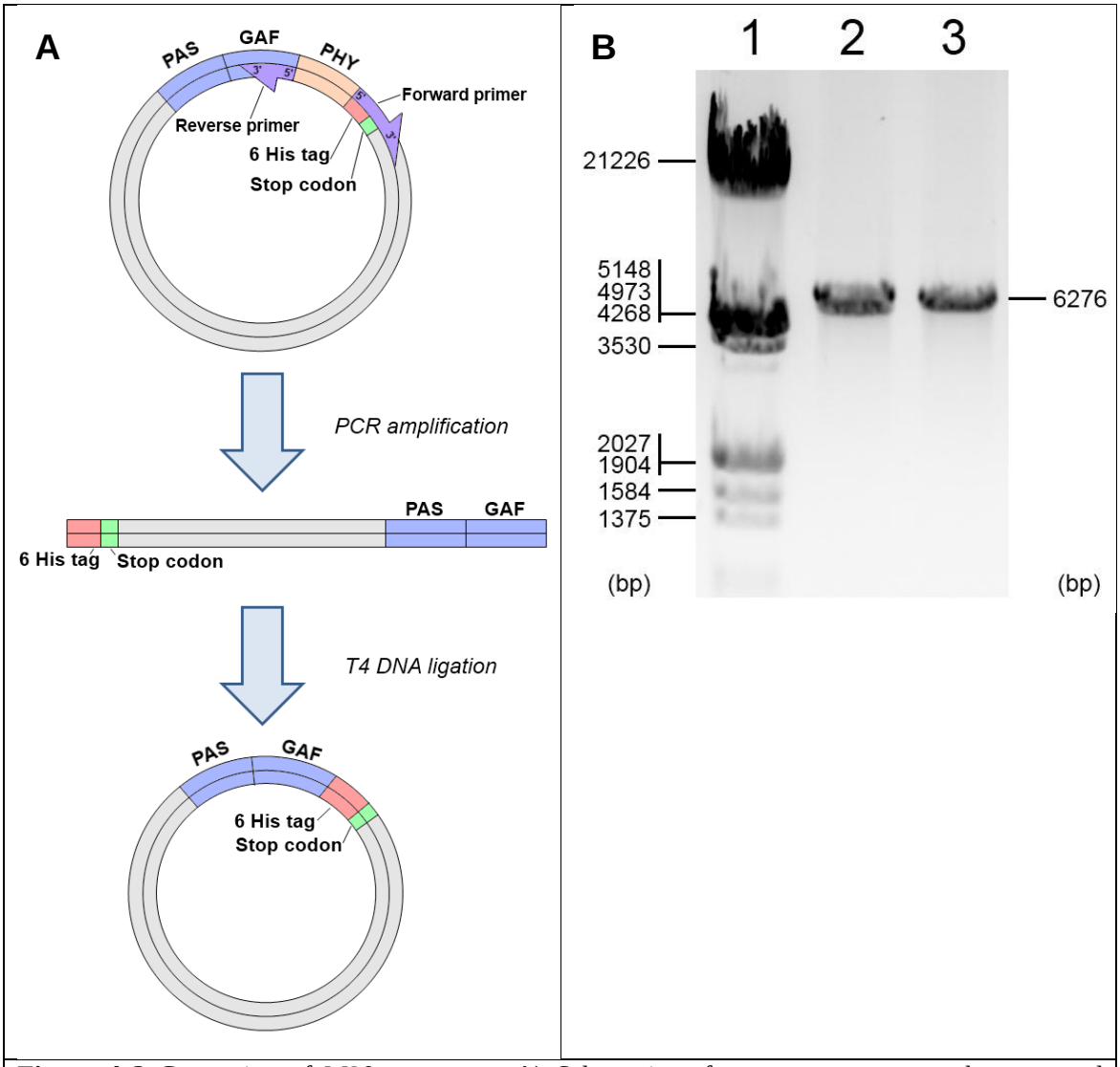


Figure A.3 Generation of M20s construct. A) Schematics of strategy to generate the truncated construct. The entire plasmid was amplified with the primers designed to flank the region to be omitted. The linear DNA fragment was ligated by blunt end ligation to create circular plasmid again by T4 DNA ligase. B) Agarose gel electrophoresis of PCR amplified DNA. Lane 1, DNA markers (bp); Lane 2 and 3, PCR amplified linear DNA that were carried out in duplication. A single band of amplicon designed to be 6276 bp is visible. The colour of the gel is inverted.

Appendix 2.2 Verification of the plasmid construct

Appendix 2.2.1 Restriction digestion

Restriction digestion was carried out to assess the success of truncation of the construct. Protein coding sequence of Agp1 PHY domain contains a restriction site for XhoI (C/TCGAG), and no other XhoI site is present in the parental pNM19a plasmid, therefore if the product plasmid resists digestion by XhoI, it implies the successful truncation. Conversely if the truncation of the plasmid is unsuccessful, XhoI digestion will yield a single linear DNA. Plasmids purified from several colonies were tested with XhoI digestion along with parental pNM19a plasmid. The restriction pattern of the XhoI treated plasmids were visualised by agarose gel electrophoresis along with non-digested plasmids (Fig. A.4), and this led to identification of several candidate plasmids as being the successfully truncated construct, M20s.

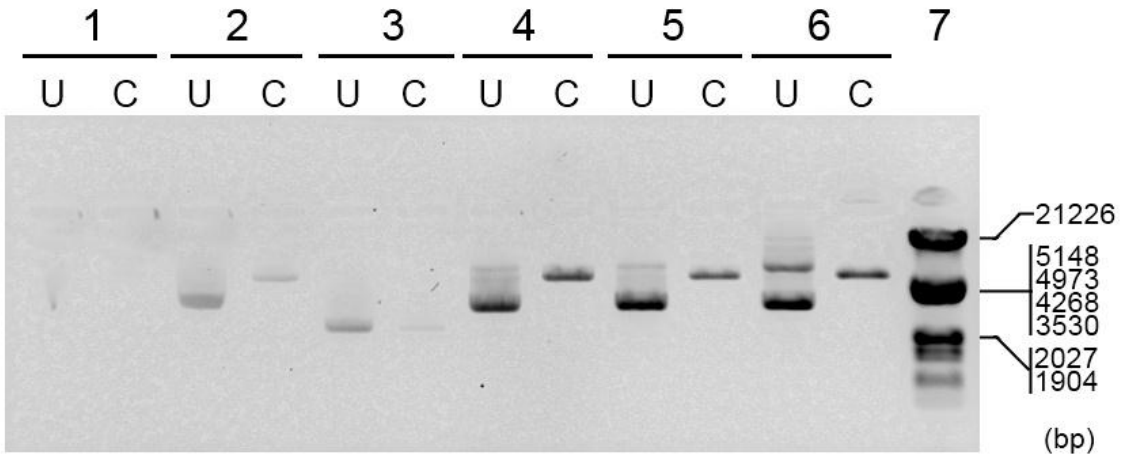


Figure A.4 Agarose gel electrophoresis of plasmid DNA before and after being subjected to XhoI digestion. Lanes 1 to 5, Candidate M20s plasmids; Lane 6, Parental pNM19a plasmid. U, Undigested with XhoI; C, Digested with XhoI. Lane 7, DNA markers (bp). Migration patterns of lane 2, 4 and 5 is similar to that of lane 6, where undigested plasmid migrates as several bands and XhoI digested plasmid appear as a single band. Only lane 3 clearly indicates a case where identical migration pattern is observed for both digested and undigested plasmid, indicating the lack of XhoI site.

Appendix 2.2.2 Test expression

Integrity of the Agp1 M20s protein coding sequence in the plasmid which exhibited positive pattern with XhoI digestion test were checked by small scale expression trials. Expression of *agp1* gene under an inducing condition will indicate the integrity of the gene and also enable to test for the success of truncation by analysing the molecular weight of the expressed protein.

E. coli BL21 (DE3) cells were transformed with three candidates of M20s construct. 10 ml LB-ampicillin cultures were treated with 100 µM IPTG. Overnight cultures were collected by centrifugation and the cells were resuspended in 2 ml of 300 mM NaCl, 50 mM Tris-HCl pH 7.8 at 20 °C, 5 mM EDTA, 14 mM 2-mercaptoethanol. The cells were disrupted by 2 cycles of freeze-thaw followed by sonication. The cell lysate was clarified by centrifugation, and soluble and insoluble fractions were analysed by SDS-PAGE (Fig. A.5). The fact that the majority of the protein was found in the insoluble fraction may be attributed to incomplete cell disruption.

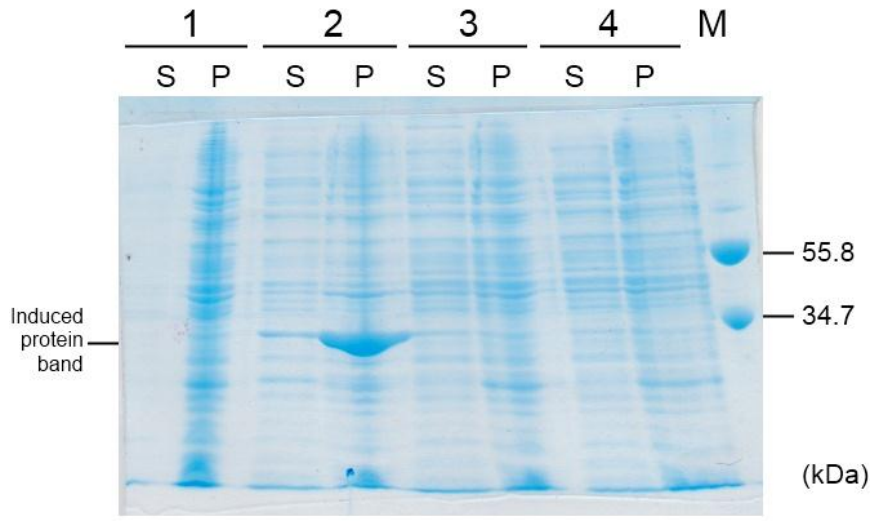


Figure A.5 SDS-PAGE gel showing small scale expression trials of M20s. BL21 (DE3) cells were transformed with of candidate plasmids for M20s, a truncated M20 construct. Groups 1 to 3, Cells after overnight induction; Group 4, Negative control where cells transformed with the same plasmid as that of lane 1 was incubated overnight without induction; M, SN molecular weight markers consisting of M15 (55.8 kDa) and M20 (34.7 kDa) apoproteins. S: Soluble fraction, P: Insoluble fraction after cell disruption. Only the candidate in lane 2 indicates the production of the M20s protein with the expected molecular weight, 33.5 kDa.

Appendix 2.2.3 DNA sequencing by Sanger's dideoxy method

The gene of Agp1 M20s was judged to be intact following a successful expression of the protein with expected molecular weight and also by the success of affinity purification, which demonstrate that the coding sequence for hexahistidine tag was in frame. Finally the correctness of the protein coding sequence was checked with dideoxy-termination sequencing. A candidate plasmid exhibiting positive results in both restriction and expression tests was sequenced by dideoxy-termination sequencing in direction from 5' to 3' using a primer designed to bind to the T7 promoter region (Eurofins).

Nucleotide sequence from the sequencing result was aligned with the reference sequence of Agp1 M20s for inspection (Fig. A.6). Many mismatches (69^{th}, 70^{th}, 71^{th}, 74^{th}, and 77^{th} positions of the sequencing result) were found in the region immediately following the start codon, but these deviations from the reference sequence were not found to cause any change in the translated amino acid sequence (Fig. A.7). The following sequencing result perfectly matched to the reference sequence apart from five apparent insertion mutations at 765^{th}, 864^{th}, 903^{rd}, 952^{nd}, and 954^{th} of the sequencing result (Fig. A.6). The apparent insertion mutations were assumed to be errors during nucleotide sequence assignment. The point mutation at 969^{th} of the sequencing result is a silent mutation even if the sequencing result is true. Thus the entire sequence of M20s (Agp1 10-313 + hexahistidine tag) was deemed correct.

```
Sequencing result  61 ATATGAGTAGCCACACACCGAAACTGGATAGTTGCGGCGCAGAGCCCATCCACATTCCCG 120
Agp1_M20_10_313    28 --ATGAGTTCACATACGCCGAAACTGGATAGTTGCGGCGCAGAGCCCATCCACATTCCCG 85
                        ******   ** ** ******************************************

Sequencing result 121 GTGCTATTCAGGAACATGGAGCCCTTCTGGTTCTGTCCGCGCGGGAATTTTCCGTGGTGC 180
Agp1_M20_10_313    86 GTGCTATTCAGGAACATGGAGCCCTTCTGGTTCTGTCCGCGCGGGAATTTTCCGTGGTGC 145
                      ************************************************************

Sequencing result 181 AGGCGAGCGATAATCTCGCCAACTATATTGGTGTGGACCTGCCAATCGGGGCCGTTGCCA 240
Agp1_M20_10_313   146 AGGCGAGCGATAATCTCGCCAACTATATTGGTGTGGACCTGCCAATCGGGGCCGTTGCCA 205
                      ************************************************************

Sequencing result 241 CCGAAGCAAACTTGCCGTTCATCTCCGTACTGTCCGCCTGGTATTCCGGCGAGGAGAGCA 300
Agp1_M20_10_313   206 CCGAAGCAAACTTGCCGTTCATCTCCGTACTGTCCGCCTGGTATTCCGGCGAGGAGAGCA 265
                      ************************************************************

Sequencing result 301 ATTTCCGATATGCCTGGGCCGAAAAGAAACTCGATGTTTCGGCGCATCGATCCGGGACCT 360
Agp1_M20_10_313   266 ATTTCCGATATGCCTGGGCCGAAAAGAAACTCGATGTTTCGGCGCATCGATCCGGGACCT 325
                      ************************************************************

Sequencing result 361 TGGTTATTCTTGAGGTGGAAAAGGCAGGAGTGGGCGAATCCGCTGAAAAGCTGATGGGCG 420
Agp1_M20_10_313   326 TGGTTATTCTTGAGGTGGAAAAGGCAGGAGTGGGCGAATCCGCTGAAAAGCTGATGGGCG 385
                      ************************************************************

Sequencing result 421 AACTCACCAGTCTGGCGAAATATCTGAACAGCGCGCCGTCGCTGGAGGACGCGCTTTTTC 480
Agp1_M20_10_313   386 AACTCACCAGTCTGGCGAAATATCTGAACAGCGCGCCGTCGCTGGAGGACGCGCTTTTTC 445
                      ************************************************************

Sequencing result 481 GGACGGCGCAACTGGTGTCCTCGATCAGCGGTCATGACCGTACCCTGATCTATGATTTCG 540
Agp1_M20_10_313   446 GGACGGCGCAACTGGTGTCCTCGATCAGCGGTCATGACCGTACCCTGATCTATGATTTCG 505
                      ************************************************************

Sequencing result 541 GACTGGACTGGAGCGGCCATGTCGTGGCGGAAGCGGGAAGCGGGGCTTTGCCGTCTTATC 600
Agp1_M20_10_313   506 GACTGGACTGGAGCGGCCATGTCGTGGCGGAAGCGGGAAGCGGGGCTTTGCCGTCTTATC 565
                      ************************************************************

Sequencing result 601 TCGGCCTTCGCTTTCCGGCGGGCGACATCCCGCCGCAGGCGCGGCAGCTCTATACCATCA 660
Agp1_M20_10_313   566 TCGGCCTTCGCTTTCCGGCGGGCGACATCCCGCCGCAGGCGCGGCAGCTCTATACCATCA 625
                      ************************************************************

Sequencing result 661 ATCGCCTGCGGATGATTCCCGATGTCGATTATAAGCCCGTGCCGATCAGACCGGAGGTGA 720
Agp1_M20_10_313   626 ATCGCCTGCGGATGATTCCCGATGTCGATTATAAGCCCGTGCCGATCAGACCGGAGGTGA 685
                      ************************************************************

Sequencing result 721 ATGCGGAAACGGGAGCGGTGCTCGATATGAGCTTTTCACAGCTTTCGCAGCGTATCGCCG 780
Agp1_M20_10_313   686 ATGCGGAAACGGGAGCGGTGCTCGATATGAGCTTTTCACAGCTT-CGCAGCGTATCGCCG 744
                      ******************************************* ***************

Sequencing result 781 GTTCATCTCGAATATATGCGCAATATGGGCACTGCCGCTTCGATGTCGGTGTCCATCGTC 840
Agp1_M20_10_313   745 GTTCATCTCGAATATATGCGCAATATGGGCACTGCCGCTTCGATGTCGGTGTCCATCGTC 804
                      ************************************************************

Sequencing result 841 GTCAATGGCGCTCTCTGGGGCTTTGATTGCCTGCCACCACGCGACACCGCATTCGGTATC 900
Agp1_M20_10_313   805 GTCAATGGCGCTCTCTGGGGCTT-GATTGCCTGCCACCACGCGACACCGCATTCGGTATC 863
                      ********************** ***********************************

Sequencing result 901 CCCTTGCCGTACGTGAAGCCTGTGATTTTGCAGCGCAGCTCCTGTCGATGCCGAATCGCG 960
Agp1_M20_10_313   864 CC-TTGCCGTACGTGAAGCCTGTGATTTTGCAGCGCAGCTCCTGTCGATGC-GGATCGCG 921
                      ** ********************************************* * ******

Sequencing result 961 ATGGAGCAAAGTTCCCAGCACCACCACCACCCACCACTGGAAATCCGGCTGCTAACAAAG 1120
Agp1_M20_10_313   922 ATGGAGCAGAGTTCCCAGCACCACCACCACCCACCAC---------------------- 947
                      ******** ****************************
```

Figure A.6 Sequencing result of M20s. Sequence alignment was carried out between reference M20s sequence and the result from the sequencing with ClustalW 2.0.9 (Larkin et al., 2007) with the default settings. Apparent mutation in the result is highlighted in blue.

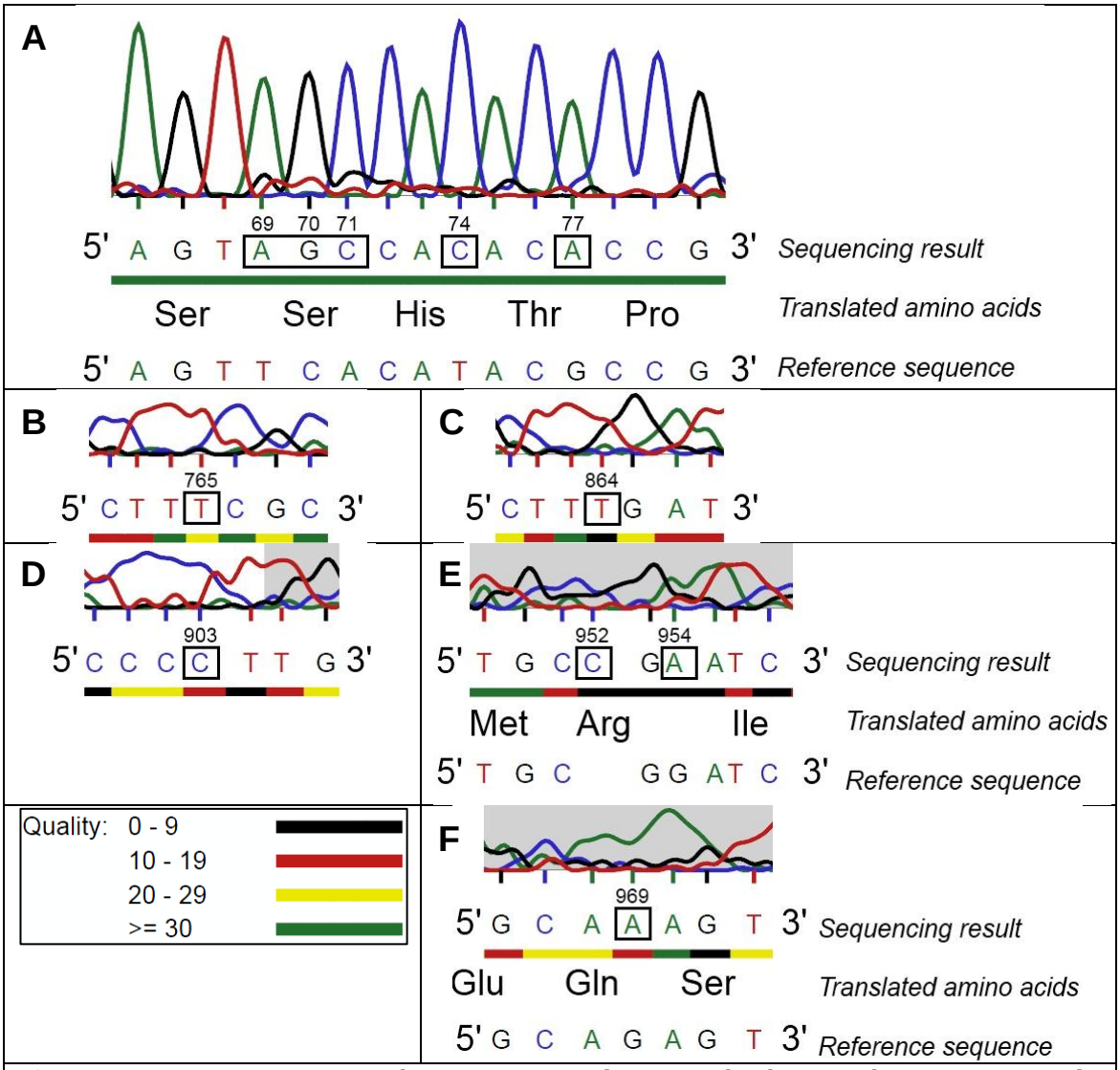


Figure A.7 Apparent mutations in the sequencing result. Mismatches between the sequencing result and the reference sequence are indicated by rectangles. A) Mismatches do not cause alteration in the translated amino acids. B) C) D) E) F) Apparent insertion mutation at 765th, 864th, 903rd, and 952nd positions are assumed to be error that had occurred during assigning sequence by the fluorescence detector, because these cases always occur on or next to a low quality region with consecutive same nucleotides. Apparent point mutations at 954th and 969th positions do not alter the translated amino acids.

Appendix 3: Generation of single cluster surface entropy reduction mutants

Appendix 3.1 QuikChange mutagenesis

Expression constructs for three mutants of Agp1-M15s were created using QuikChange site directed mutagenesis protocol (Agilent technologies). The mutagenesis reactions were carried out in PCR thermal cycler with pNM19a as the template DNA and the forward and the reverse primers for the corresponding mutation as indicated in Section 2.2.2.9. Successful amplification of the plasmids were confirmed by agarose gel electrophoresis (Fig. A.8). Plasmids were purified from several *E. coli* XL1-Blue colonies bearing mutated pNM19a plasmids, and the candidate plasmids were verified for successful mutations as described below.

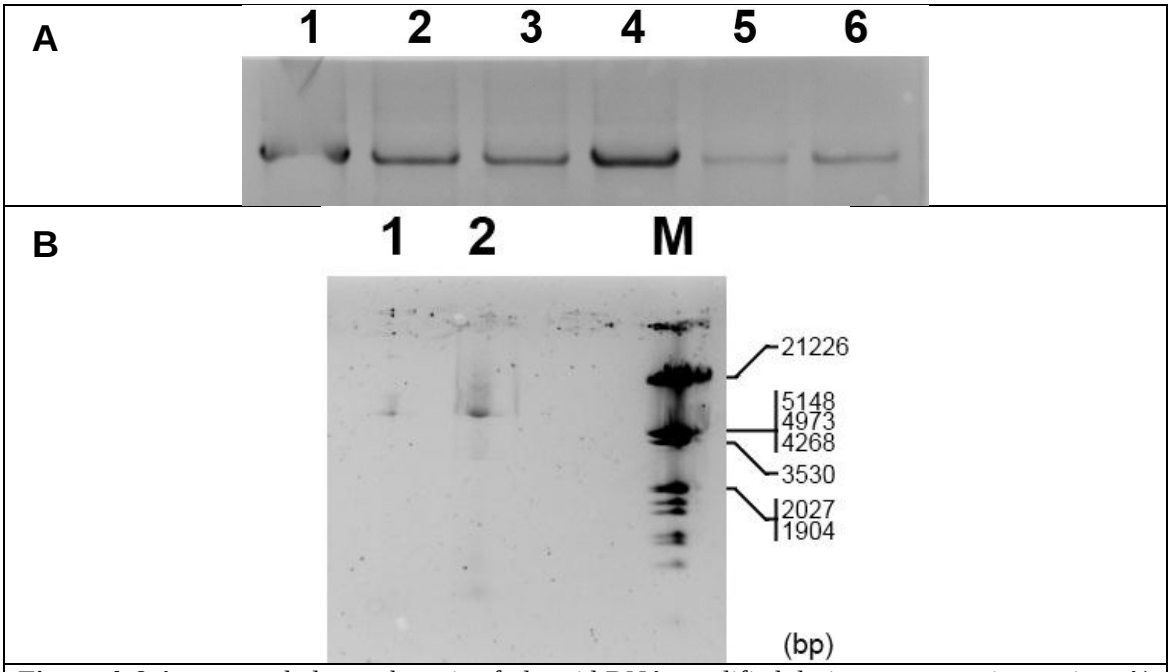


Figure A.8 Agarose gel electrophoresis of plasmid DNA amplified during mutagenesis reaction. A) DNA amplified during SER3 mutagenesis. Lane 1, 50 ng template + 3% DMSO; Lane 2, 50 ng template + 1.5 mM $MgCl_2$; Lane 3, 50 ng template + 3% DMSO + 1.5 mM $MgCl_2$; Lane 4, 10 ng template + 3% DMSO; Lane 5, 10 ng template + 1.5 mM $MgCl_2$; Lane 6, 10 ng template + 3% DMSO + 1.5 mM $MgCl_2$. B) DNA amplified during SER2 and SER6 mutagenesis. Lane 1, SER2; Lane 2, SER6; M, DNA sizer III marker. Faint bands are visible for both SER2 and SER6 between 21226 bp and 5148 bp. Colour of the gels are inverted for the visibility purpose.

Appendix 3.2 Verification of the plasmid constructs

Appendix 3.2.1 Restriction digestion

Successful mutations were first verified with restriction digestion. As well as the mutating the target residues into Alanines, the recognition sites of the following restriction enzymes were incorporated into the plasmids, and were used as a marker for successful mutagenesis - SER3: BglI (Fig. A9), SER2: NheI, and SER6: BglI (Fig. A.10).

For SER3 test groups, all cases of undigested plasmids migrated as multiple bands indicating the numbers of alternative coiling state of the plasmids. Restriction pattern of candidate number 3 and 4 are similar to that of the WT plasmid that two bands are present at the expected fragment sizes. Restriction patterns of candidate number 1, 2, 5, and 6 were in accordance with expected fragment sizes of BglI digested SER3 plasmids. The upper band was assigned to be 2654 bp fragment while the lower band comprises both 2050 bp and 2151 bp fragments with similar migration rate on a gel at the resolving capacity of 0.7% agarose gel.

For SER2 test group, the migration pattern of the parental plasmid was identical for both undigested plasmid and the plasmid subjected to NheI digestion. This indicates the lack of NheI restriction site in the WT plasmid as expected. All three candidates digested by NheI migrated as a single band at the expected size of 6855 bp, indicating successful SER2 mutagenesis.

All three candidates digested with BglI in SER6 test group exhibited restriction pattern distinct from that of the WT plasmid. The upper band was assigned to be 2868 bp fragment while the lower band comprises both 2050 bp and 1937 bp fragments with similar migration rate on a gel at the resolving capacity of 0.7% agarose gel.

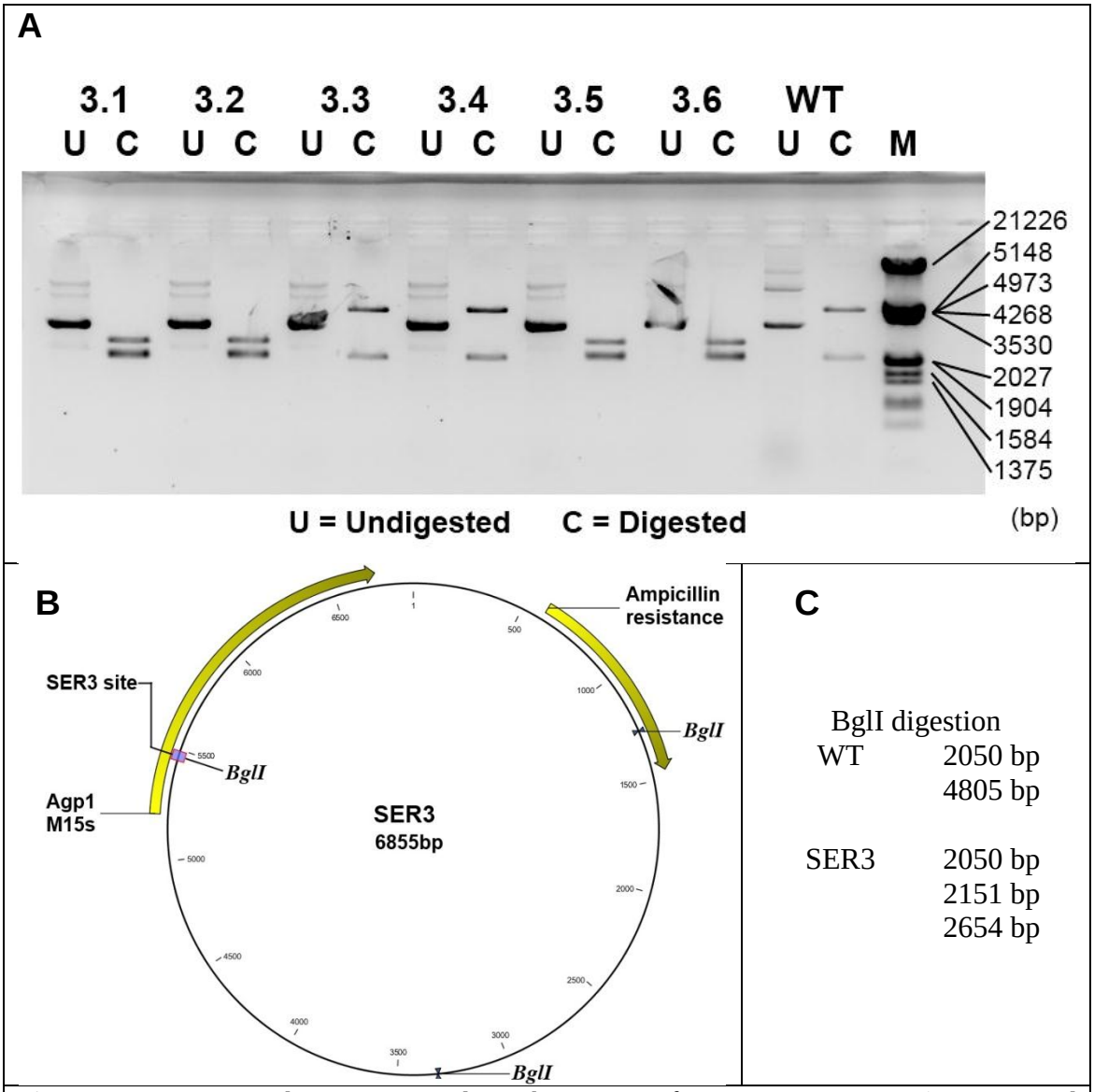


Figure A.9 Restriction digestions to evaluate the success of SER3 mutagenesis. A) Agarose gel electrophoresis of BglI restriction digested plasmids (C) and undigested plasmids (U) carried out with a 0.7% agarose gel. The colour of the gel is inverted for the visibility purpose. Plasmids from six colonies tested for successful SER3 mutagenesis is represented by the number after the decimal point. WT: Parental WT plasmid was also included as the negative control. M: Molecular weight marker (bp). B) Schematic diagram of the parental plasmid with the BglI sites indicated. SER3 site is also shown, where an additional BglI site is incorporated. C) Expected sizes of the fragments of WT and SER3 plasmids following BglI digestion.

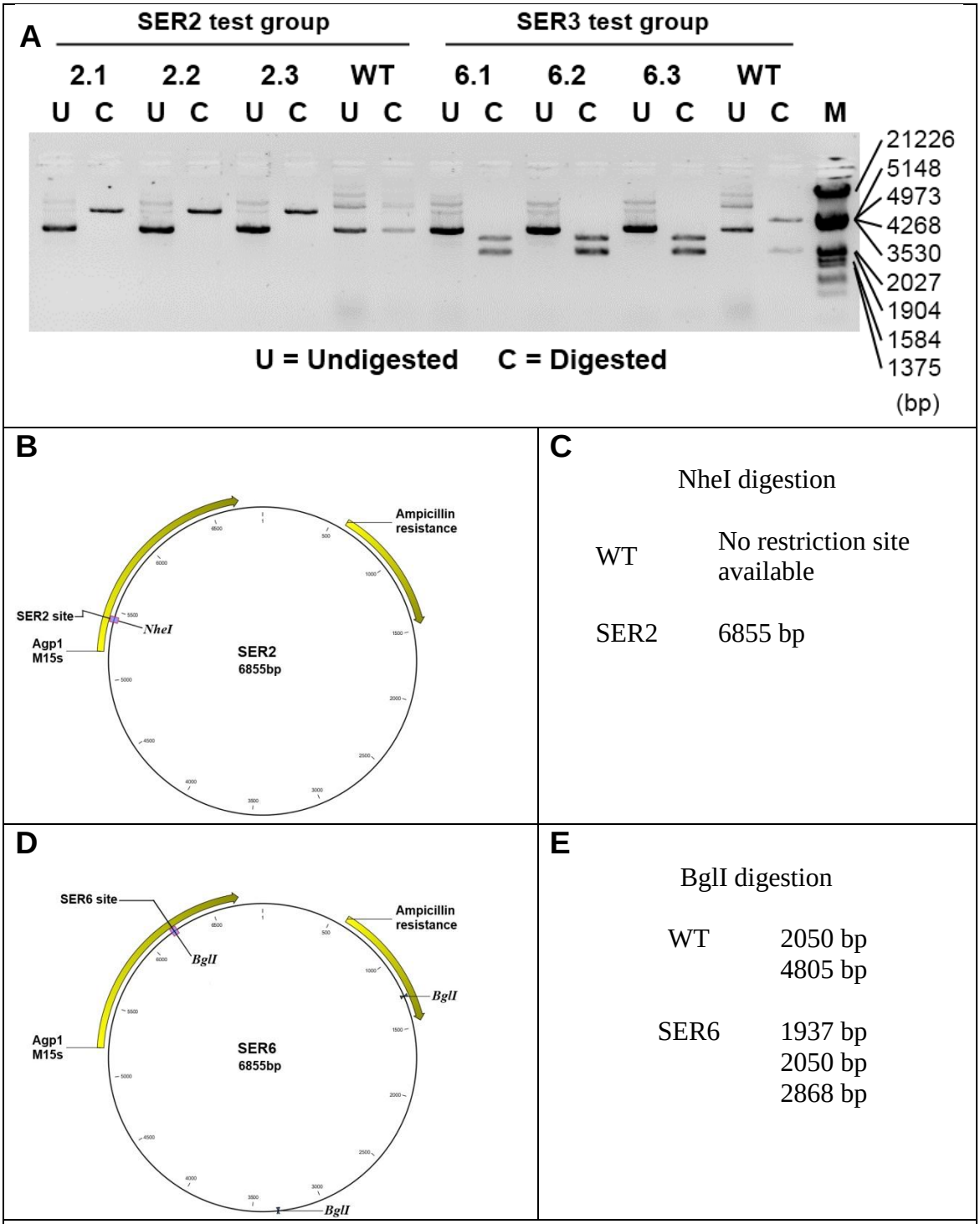


Figure A.10 Restriction digestions to evaluate the success of SER2 and SER6 mutagenesis. A) Plasmids digested with restriction enzymes and undigested plasmids were separated by 0.7% agarose gel electrophoresis. NheI and BglI were used for identifying successful SER2 and SER6 mutagenesis, respectively. The colour in inverted for the purpose of visibility. Plasmids from three colonies were tested for successful SER2 and SER6 mutagenesis each, and are represented by the number after the decimal point. Schematic diagram of the plasmid and the locations of the restrictions sites. B) A NheI site is introduced by SER2 mutation. C) Expected sizes of the fragments of WT and SER2 plasmids following NheI digestion. D) A BglI site is introduced by SER6 mutation. E) Expected sizes of the fragments of WT and SER6 plasmids following BglI digestion.

Appendix 3.2.2 Test expression

For those plasmids indicated to be successfully mutated by restriction digestion were tested for the integrity of the Agp1 sequence and ability to express protein. Purified plasmids were transformed into *E. coli* BL21(DE3) cells and small scale test expressions were carried out.

Each SER plasmid candidate was subjected to small scale test expression with 5 ml of LB-Amp media according to the Section 2.2.3.6.1. The soluble lysate and the insoluble pellet of each candidate was analysed by SDS-PAGE for each candidate (Fig. A.11).

For SER2 mutation only candidate number 2 successfully resulted in production of the target protein while four and two candidates resulted in production for mutations SER3 and SER6, respectively. Molecular weights of the overproduced proteins form the successful candidates were the same as that of the WT M15 protein. In majority of the cases, the overproduced proteins were only identified in the insoluble fraction of lysate, even in the case of expression of WT M15 protein. Unpredictable presence of the target protein in the soluble fraction was attributed to incomplete cell lysis.

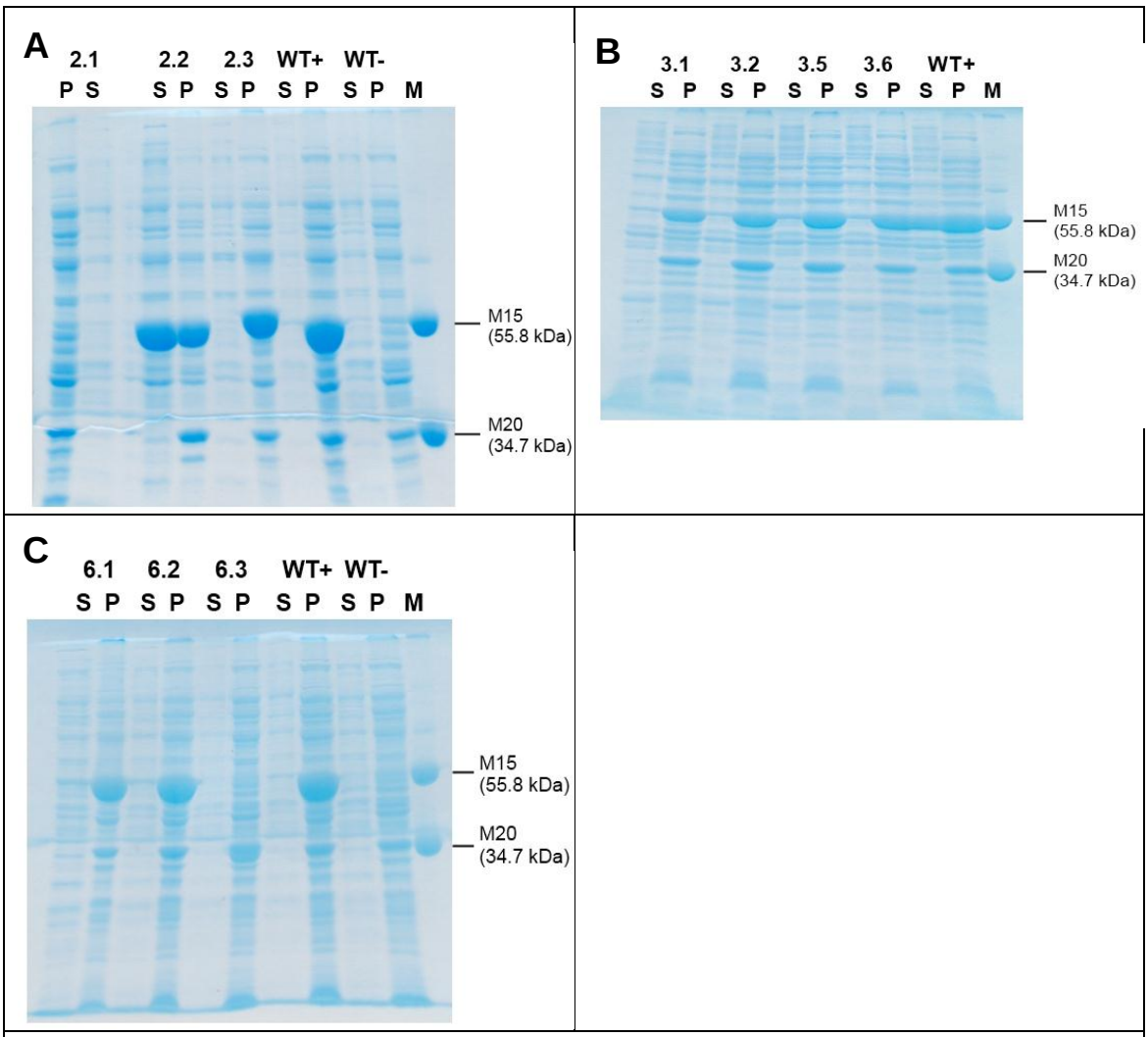


Figure A.11 SDS-PAGE gel showing the small scale expression trials with the SER mutant candidate plasmids. Cells bearing WT plasmid was also tested as the control. A) No indication of the SER2 protein overproduction is present in the candidate 1. SDS-PAGE of SER2 test group. Candidate 2 produces protein of the same molecular weight as the WT protein. Molecular weight of the protein produced by candidate 3 appears larger than the WT. B) All 4 SER3 candidates produced protein of the same molecular weight as the WT protein. C) SDS-PAGE of SER6 test group. Candidate 1 and 2 produced protein of the same molecular weight as the WT protein, while no protein overproduction was evident for candidate 3. M, molecular weight marker consisting of purified M15 apoprotein and M20 apoprotein. WT+, WT induced; WT-, WT uninduced.

Appendix 3.2.3 DNA sequencing by Sanger's dideoxy method

Successful mutagenesis was finally verified with DNA sequencing. Plasmid candidates SER2.2, SER3.5, and SER6.2 were selected for sequencing with T7 forward primer for SER2 and SER3, while T7 reverse primer was selected for SER6 (Section 2.2.2.10). Successful mutagenesis was confirmed for all SER mutants under high quality regions (Fig. A.12).

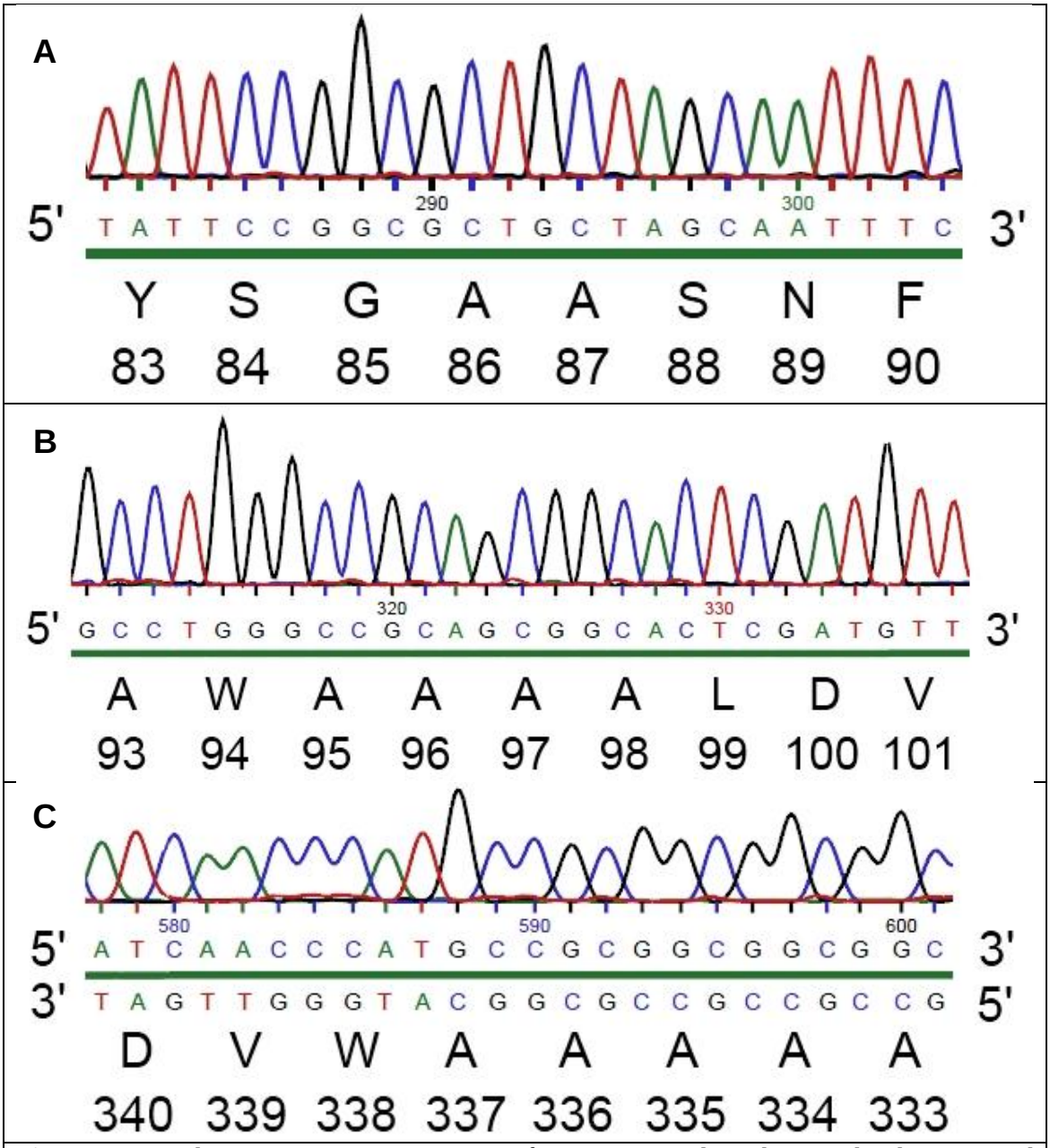


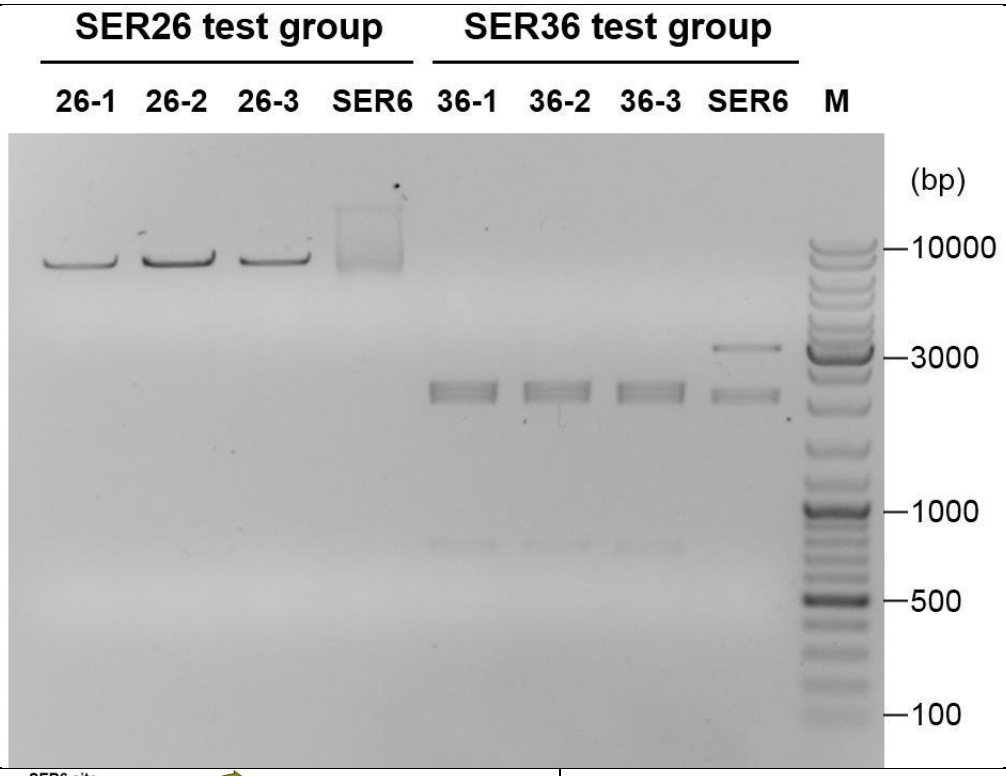
Figure A.12 Dideoxy-termination sequencing of SER mutant plasmids. Translated amino acid sequences in the correct reading frame are annotated with the residue number. A) SER2 (E86A/E87A). B) SER3 (E96A/K97A/K98A). C) SER6 (E336A/K337A). T7 reverse primer was used for sequencing this mutant. Complementally nucleotide sequence is also shown to ease the identification of the translated amino acid sequence, shown in reverse order.

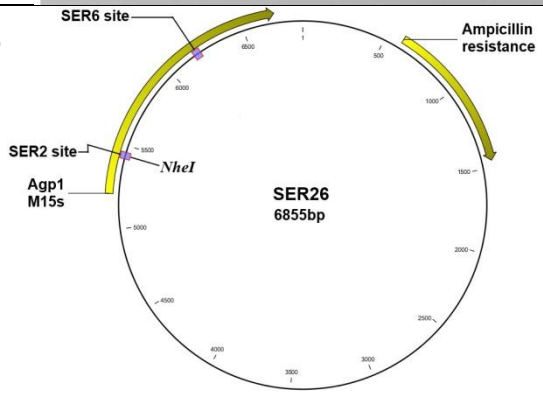
Appendix 4: Sequencing result of double cluster SER mutants

Appendix 4.1 Verification of the plasmid constructs

Appendix 4.1.1 Restriction digestion

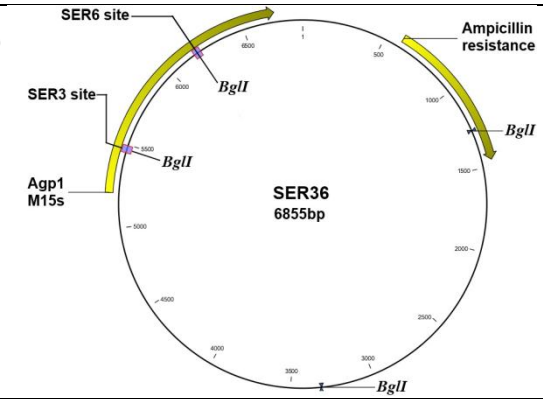
Successful mutagenesis was firstly tested by restriction digestion. Candidates for SER26 were digested by NheI to confirm the successful introduction of SER2 mutation to parental the SER6 plasmid, which like WT pNM19a doesn't contain any NheI site. Candidates for SER36 were digested by BglI to confirm the successful introduction of SER3 mutation to the parental SER6 plasmid. SER3 mutation introduces an additional BglI site to SER6 plasmid which has 3 BglI sites, of which one of them is as a result of SER6 mutation. All candidates for both mutations indicated successful mutagenesis exhibiting expended band sizes on an agarose gel (Fig. A.13).





C

NheI digestion	
SER6	No restriction site available
SER26	6855 bp



E

BglI digestion	
SER6	1937 bp
	2050 bp
	2868 bp
SER36	2151 bp
	2050 bp
	1937 bp
	717 bp

Figure A.13 Restriction digestions to evaluate the success of SER26 and SER26 mutagenesis. A) Agarose gel electrophoresis of plasmids digested with restriction enzyme. The colour of the gel has been inverted.

Appendix 4.1.2 Test expression

Test expression was carried out on the constructs with positive sign of mutagenesis to test for the integrity of the protein coding sequence. Based on an SDS-PAGE gel (Fig. A.14), the constructs SER26.2 and SER36.3 were selected for further verification.

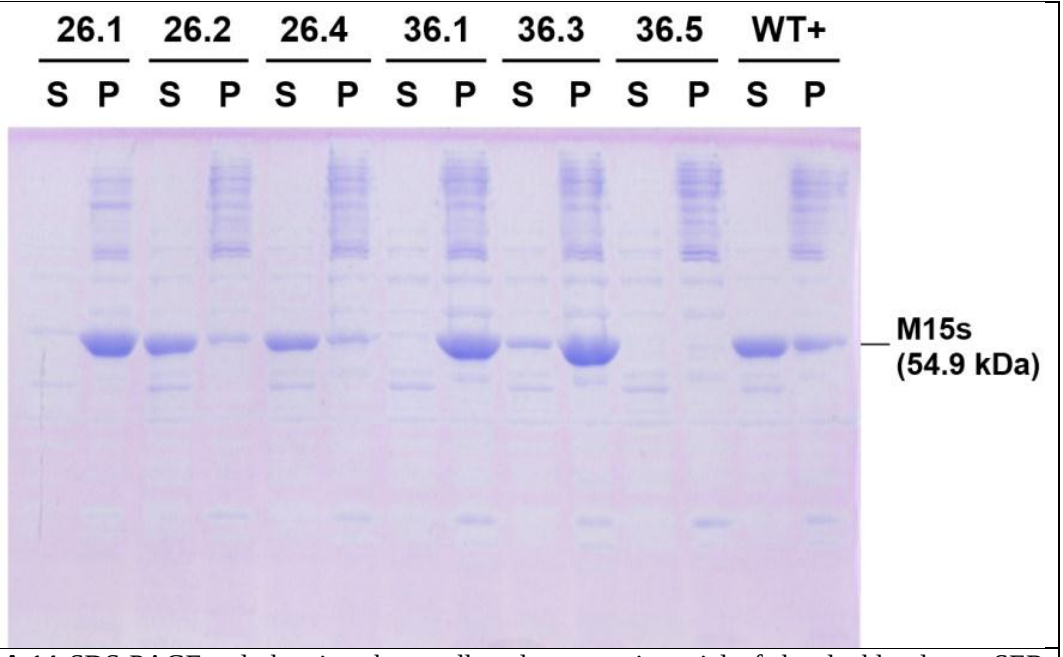


Figure A.14 SDS-PAGE gel showing the small scale expression trial of the double cluster SER candidate plasmids. S, Soluble fraction; P, Insoluble fraction of the cell lysate. WT+, positive control with the construct for WT M15s protein.

Appendix 4.1.3 DNA sequencing by Sanger's dideoxy method

Successful mutagenesis was finally verified with DNA sequencing. Plasmid candidates SER26.2 and SER36.3 were selected for sequencing with T7 forward primer for both SER26 and SER36 (Fig. A15). Verification of the SER6 mutation was deemed unnecessary since the mutagenesis of SER26 and SER36 plasmids were based on SER6.2, whose sequence has been verified (Appendix 3.2.3).

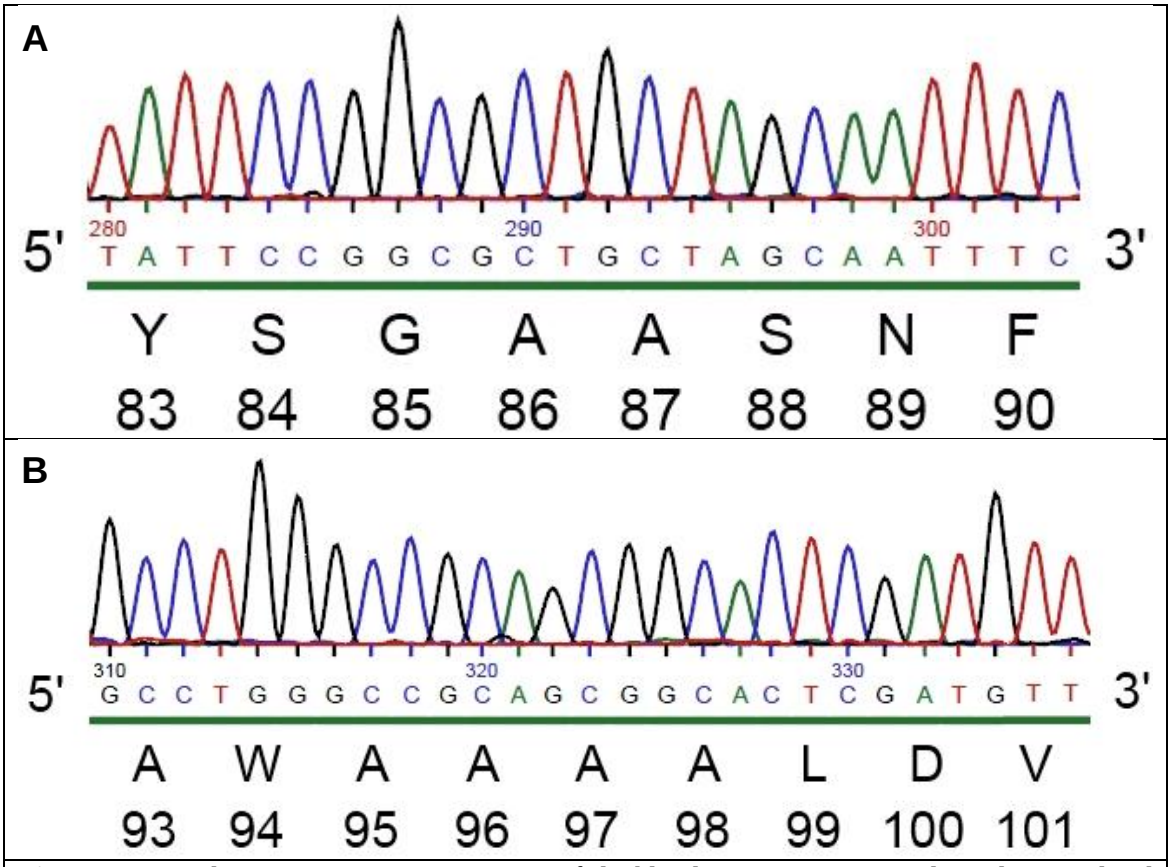


Figure A.15 Dideoxy-termination sequencing of double cluster SER mutant plasmids. Translated amino acid sequences in the correct reading frame are annotated with the residue number. Since a successfully mutated SER6 was used as the template for generation of both SER26 and SER36, only unique mutation sites are shown here. A) SER26 (E86A/E87A/E336A/K337A). B) SER36 (E96A/K97A/K98A/E336A/K337A).

Appendix 5: Calibration of Superdex 200 column

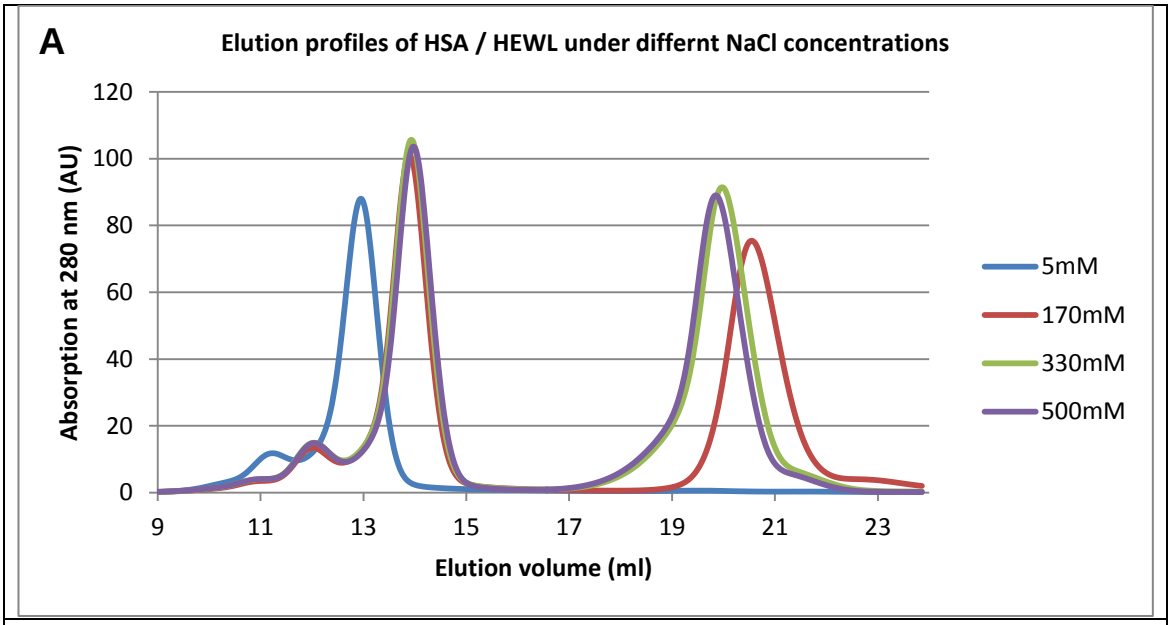
To enable correlation between the elution volumes and the molecular weight of the eluate, calibration was carried out with two model proteins, human serum albumin (HSA) and hen egg white lysozyme (HEWL) (Table A.1).

Calibration of the Superdex 200 column was carried out for buffers at several NaCl concentrations. Superdex 200 column was equilibrated with SEC buffer (20 mM Tris-HCl pH 7.8 at 20°C, 1 mM EDTA) containing either 500, 330, 170, or 5 mM of NaCl. 100 µl of protein solutions containing 450 mAU×ml of each protein were prepared in SEC buffer of the same NaCl concentration as the buffer used in the calibration run. The mixtures were subjected to isocratic elution with S200 column, and the chromatograms are shown in Fig. A.16.

HSA eluted in two peaks, and the elution volumes increased with increasing NaCl concentration. HEWL eluted in a single peak, and the elution volume decreased with increasing NaCl concentration. NaCl concentration dependent mobility of the elution peak indicates slight degree of hydrophobic interaction and ionic interaction with the stationary phase for HSA and HEWL, respectively. The first minor peak of HSA was attributed to be its dimer as judged by SDS-PAGE gel (data not shown) and the by the calibration curve drawn by other two peaks. Majority of HEWL failed to elute with eluent containing 5 mM NaCl, likely due to non-specific interaction with the stationary phase. A very minor peak, indiscernible at the scale of presentation in Fig. A.16 was detected at 19.68 ml, but since this value is against the trend of decreasing elution volume with increasing NaCl concentration, it was omitted as an anomaly, and instead the elution volume of HEWL at 170 mM NaCl was used for calculating the calibration curve for 5 mM (Fig. A.17).

Protein	**Human serum albumin (HSA)**	**Hen egg white lysozyme (HEWL)**
Source	Sigma A3782-5G	Fluka 62970
MW specified by the supplier	66478 Da / monomer	14300 Da / monomer
Seq. accession number used	AAA98798	ACL81750.1
The following was calculated by Protein calculator		
MW (Da)	69366.4844	16162.5049
pI	6.29	9.09
# of Cys (# of disulphide bonds)	35 (17)	9 (4)
ε280 with disulphide bonds	37740	37180

Table A.1 Biochemical information of two model proteins used for calibration of Superdex 200 column.

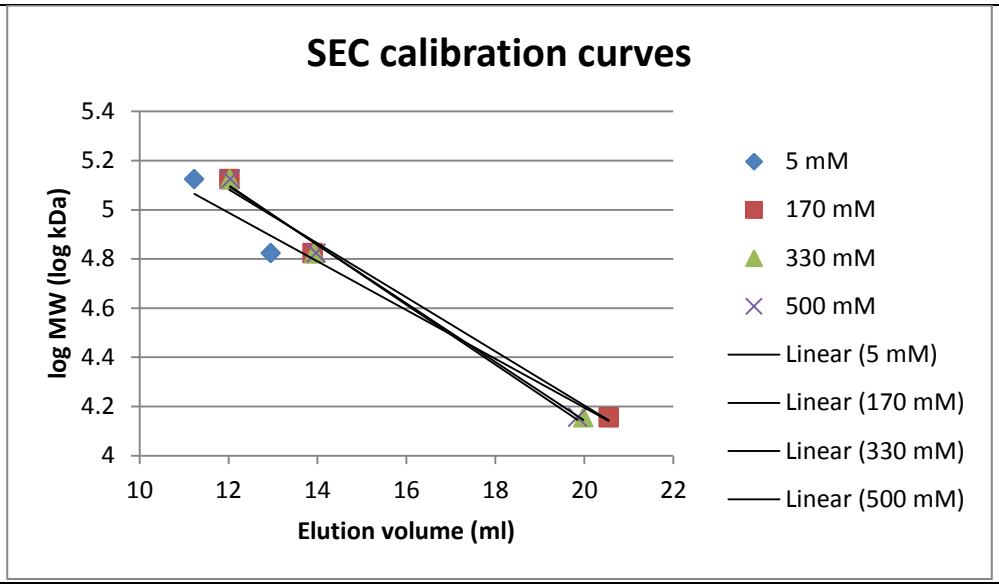


B

		NaCl concentration in the eluent			
	Identity	*5 mM*	*170 mM*	*330 mM*	*500 mM*
Peak 1	HSA dimer	11.23	12.02	12.03	12.04
Peak 2	HSA monomer	12.95	13.89	13.93	13.97
Peak 3	HEWL monomer	19.68	20.55	19.98	19.85

Figure A.16 Chromatograms from calibration of Superdex 200 SEC column. A) Chromatograms of HSA and HEWL in eluent under various NaCl concentrations. B) Elution volumes of the peaks observed during the Superdex 200 calibration. Units of the elution volumes are in ml. The value for HEWL at 5 mM NaCl is indicated by grey highlight. It was omitted as an anomaly against the trend, and was replaced by the value for 170 mM for calculating the calibration curve for 5 mM. Chromatograms were obtained by monitoring absorption at 280 nm.

A



B

5 mM NaCl

$$y = -0.099\,x + 6.177$$

170 mM NaCl

$$y = -0.110\,x + 6.404$$

330 mM NaCl

$$y = -0.119\,x + 6.523$$

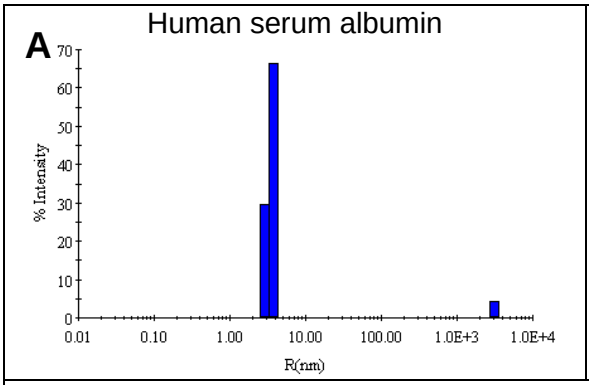
500 mM NaCl

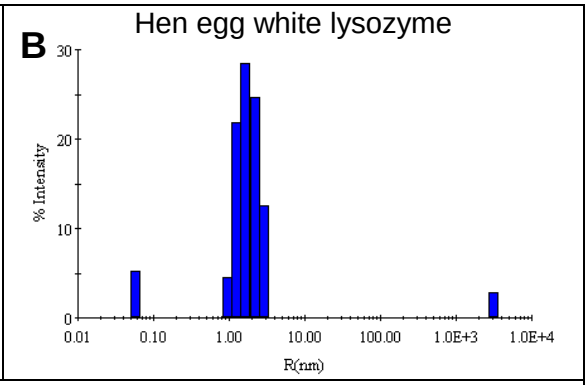
$$y = -0.122\,x + 6.559$$

Figure A.17 Calibration curves from calibration of Superdex 200 SEC column. A) Calibration curves with elution volumes (x-axis) and log of molecular weight of eluates (y-axis). B) Calculated calibration curves for each NaCl concentration, which were used for estimating the molecular weight of the eluate during analytical SEC. y = log of molecular weight of the eluate, and x = elution volume (ml).

Appendix 6: Calibration of dynamic light scattering

Accuracy of the DLS was verified by taking measurements of two model proteins, HSA and HEWL. Solution of each protein was prepared at 1.0 mg/ml in SEC buffer with 170 mM NaCl, and 20 measurements were taken from which average peaks were obtained (Figure A.18). There were good agreement between the molecular weight information supplied by the manufacturer and the calculated molecular weight from DLS, thereby verifying the use of DLS to probe the apparent molecular weight of the protein in solution.





C

Protein type, [NaCl], [Protein]		R (nm)	%Pd	MW (kDa)	%Int	%Mass
Human Serum Albumin	Peak 1	3.589	12.2	67	95.8	100
170 mM, 1.0 mg/ml	Peak 2	3195.590	0	533491000	4.2	0
Hen Egg White Lysozyme	Peak 1	0.058	0	0	5.2	99.9
170mM 1.0 mg/ml	Peak 2	1.860	30.6	14	92	0.1
	Peak 3	3195.590	0	533491000	2.8	0

Figure A.18 Verification of DLS measurements. A) HSA and B) HEWL were subjected to DLS, which returned the values similar to the molecular weight of the samples (HSA 66478 Da and HEWL 14300 Da). C) Tabulated results. R, radius; Pd, polydispercity; MW, molecular weight; Int, intensity. The peak with the highest % intensity value is highlight in grey. Values in the table are when the plots are made with % Intensity as the Y-axis and the solvent is set as PBS.

References

ABRAMOFF, M. D., MAGALHÃES, P. J. & RAM, S. J. 2004. Image processing with ImageJ. *Biophotonics international,* 11**,** 36-42.

ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L.-W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D,* 66**,** 213-221.

AULDRIDGE, M. E., SATYSHUR, K. A., ANSTROM, D. M. & FOREST, K. T. 2012. Structure-guided engineering enhances a phytochrome-based infrared fluorescent protein. *The Journal of Biological Chemistry,* 287**,** 7000-7009.

AVBELJ, F. & FELE, L. 1998. Role of main-chain electrostatics, hydrophobic effect and side-chain conformational entropy in determining the secondary structure of proteins. *Journal of Molecular Biology,* 279**,** 665-684.

BAILEY, S. 1994. The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D,* 50**,** 760-763.

BAKER, N. A., SEPT, D., JOSEPH, S., HOLST, M. J. & MCCAMMON, J. A. 2001. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 10037-10041.

BATTYE, T. G., KONTOGIANNIS, L., JOHNSON, O., POWELL, H. R. & LESLIE, A. G. 2011. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D,* 67**,** 271-281.

BAUMGARTNER, J. W., KIM, C., BRISSETTE, R. E., INOUYE, M., PARK, C. & HAZELBAUER, G. L. 1994. Transmembrane signalling by a hybrid protein: communication from the domain of chemoreceptor Trg that recognizes sugar-binding proteins to the kinase/phosphatase domain of osmosensor EnvZ. *Journal of Bacteriology,* 176**,** 1157-1163.

BELLINI, D. & PAPIZ, M. Z. 2012a. Dimerization properties of the RpBphP2 chromophore-binding domain crystallized by homologue-directed mutagenesis. *Acta Crystallographica Section D,* 68**,** 1058-1066.

BELLINI, D. & PAPIZ, MIROSLAV Z. 2012b. Structure of a Bacteriophytochrome and Light-Stimulated Protomer Swapping with a Gene Repressor. *Structure (London, England : 1993),* 20**,** 1436-1446.

BENDA, C., SCHEUFLER, C., DE MARSAC, N. T. & GÄRTNER, W. 2004. Crystal Structures of Two Cyanobacterial Response Regulators in Apo- and Phosphorylated Form Reveal a Novel Dimerization Motif of Phytochrome-Associated Response Regulators. *Biophysical Journal,* 87**,** 476-487.

BENVENUTI, M. & MANGANI, S. 2007. Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nature Protocols,* 2**,** 1633-1651.

BERGFORS, T. M. 2009. *Protein Crystallization,* International University Line.

BICK, M. J., LAMOUR, V., RAJASHANKAR, K. R., GORDIYENKO, Y., ROBINSON, C. V. & DARST, S. A. 2009. How to Switch Off a Histidine Kinase: Crystal Structure of Geobacillus stearothermophilus KinB with the inhibitor Sda. *Journal of Molecular Biology,* 386**,** 163-177.

BILWES, A. M., ALEX, L. A., CRANE, B. R. & SIMON, M. I. 1999. Structure of CheA, a signal-transducing histidine kinase. *Cell,* 96**,** 131-141.

BILWES, A. M., QUEZADA, C. M., CROAL, L. R., CRANE, B. R. & SIMON, M. I. 2001. Nucleotide binding by the histidine kinase CheA. *Nature Structural Biology,* 8**,** 353-360.

BLATTNER, F. R. 1997. The Complete Genome Sequence of Escherichia coli K-12. *Science,* 277**,** 1453-1462.

BLOW, D. 2002. *Outline of Crystallography for Biologists,* Oxford University Press.

BLUMENSTEIN, A., VIENKEN, K., TASLER, R., PURSCHWITZ, J., VEITH, D., FRANKENBERG-DINKEL, N. & FISCHER, R. 2005. The Aspergillus nidulans

phytochrome FphA represses sexual development in red light. *Current biology : CB,* 15**,** 1833-1838.

BLUNDELL, T. & JOHNSON, L. 1976. *Protein crystallography*, Academic Press New York:.

BORUCKI, B., OTTO, H., ROTTWINKEL, G., HUGHES, J., HEYN, M. P. & LAMPARTER, T. 2003. Mechanism of Cph1 phytochrome assembly from stopped-flow kinetics and circular dichroism. *Biochemistry,* 42**,** 13684-13697.

BORUCKI, B., VON STETTEN, D., SEIBECK, S., LAMPARTER, T., MICHAEL, N., MROGINSKI, M. A., OTTO, H., MURGIDA, D. H., HEYN, M. P. & HILDEBRANDT, P. 2005. Light-induced proton release of phytochrome is coupled to the transient deprotonation of the tetrapyrrole chromophore. *The Journal of Biological Chemistry,* 280**,** 34358-34364.

BOURRET, R. B. & SILVERSMITH, R. E. 2010. Two-component signal transduction. *Current Opinion in Microbiology,* 13**,** 113-115.

BRAATSCH, S., JOHNSON, J. A., NOLL, K. & BEATTY, J. T. 2007. The O2-responsive repressor PpsR2 but not PpsR1 transduces a light signal sensed by the BphP1 phytochrome in Rhodopseudomonas palustris CGA009. *FEMS Microbiology Letters,* 272**,** 60-64.

BUTLER, W. L., NORRIS, K. H., SIEGELMAN, H. W. & HENDRICKS, S. B. 1959. Detection, Assay, and Preliminary Purification of the Pigment Controlling Photoresponsive Development of Plants. *Proceedings of the National Academy of Sciences of the United States of America,* 45**,** 1703-1708.

CAPRA, E. J., PERCHUK, B. S., LUBIN, E. A., ASHENBERG, O., SKERKER, J. M. & LAUB, M. T. 2010. Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways. *PLoS genetics,* 6**,** e1001220.

CASINO, P., RUBIO, V. & MARINA, A. 2009. Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction. *Cell,* 139**,** 325-336.

CHEN, M., CHORY, J. & FANKHAUSER, C. 2004. Light signal transduction in higher plants. *Annual Review of Genetics,* 38**,** 87-117.

CHEUNG, J. & HENDRICKSON, W. A. 2010. Sensor domains of two-component regulatory systems. *Current Opinion in Microbiology,* 13**,** 116-123.

CHRISTOU, P. 1996. Transformation technology. *Trends in Plant Science,* 1**,** 423-431.

CIERPICKI, T., KIM, M. H., COOPER, D. R., DEREWENDA, U., BUSHWELLER, J. H. & DEREWENDA, Z. S. 2006. The DC-module of doublecortin: dynamics, domain boundaries, and functional implications. *Proteins,* 64**,** 874-882.

COOPER, D. R., BOCZEK, T., GRELEWSKA, K., PINKOWSKA, M., SIKORSKA, M., ZAWADZKI, M. & DEREWENDA, Z. 2007. Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallographica Section D,* 63**,** 636-645.

CORNILESCU, G., ULIJASZ, A. T., CORNILESCU, C. C., MARKLEY, J. L. & VIERSTRA, R. D. 2008. Solution Structure of a Cyanobacterial Phytochrome GAF Domain in the Red-Light-Absorbing Ground State. *Journal of Molecular Biology,* 383**,** 403-413.

CZEPAS, J., DEVEDJIEV, Y., KROWARSCH, D., DEREWENDA, U., OTLEWSKI, J. & DEREWENDA, Z. S. 2004. The impact of Lys-->Arg surface mutations on the crystallization of the globular domain of RhoGDI. *Acta Crystallographica Section D,* 60**,** 275-280.

DASGUPTA, S., IYER, G. H., BRYANT, S. H., LAWRENCE, C. E. & BELL, J. A. 1997. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins,* 28**,** 494-514.

DAUTER, Z., BOTOS, I., LARONDE-LEBLANC, N. & WLODAWER, A. 2005. Pathological crystallography: case studies of several unusual macromolecular crystals. *Acta Crystallographica Section D,* 61**,** 967-975.

DAVIS, S. J. 1999. Bacteriophytochromes: Phytochrome-Like Photoreceptors from Nonphotosynthetic Eubacteria. *Science,* 286**,** 2517-2520.

DELANO, W. L. 2002. *The PyMOL Molecular Graphics System* [Online]. Available: http://www.pymol.org.

DEREWENDA, Z. S. 2004a. Rational Protein Crystallization by Mutational Surface Engineering. *Structure,* 12**,** 529-535.

DEREWENDA, Z. S. 2004b. The use of recombinant methods and molecular engineering in protein crystallization. *Methods,* 34**,** 354-363.

DEREWENDA, Z. S. 2011. It's all in the crystals. *Acta Crystallographica Section D,* 67**,** 243-248.

DEREWENDA, Z. S. & VEKILOV, P. G. 2006. Entropy and surface engineering in protein crystallization. *Acta Crystallographica Section D,* 62**,** 116-124.

DIEDERICHS, K. & KARPLUS, P. A. 1997. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Structural Biology,* 4**,** 269-275.

DOBRINDT, U. & HACKER, J. 2001. Whole genome plasticity in pathogenic bacteria. *Current Opinion in Microbiology,* 4**,** 550-557.

DOLINSKY, T. J., CZODROWSKI, P., LI, H., NIELSEN, J. E., JENSEN, J. H., KLEBE, G. & BAKER, N. A. 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research,* 35**,** W522-W525.

DOLOMANOV, O. V., BOURHIS, L. J., GILDEA, R. J., HOWARD, J. A. K. & PUSCHMANN, H. 2009. OLEX2: a complete structure solution, refinement and analysis program. *Journal of Applied Crystallography,* 42**,** 339-341.

DONG, A., XU, X. & EDWARDS, A. M. 2007. In situ proteolysis for protein crystallization and structure determination. *Nature Methods,* 4**,** 1019-1021.

DONGRE, M., SINGH, N. S., DUREJA, C., PEDDADA, N., SOLANKI, A. K., ASHISH & RAYCHAUDHURI, S. 2011. Evidence on How a Conserved Glycine in the Hinge Region of HapR Regulates Its DNA Binding Ability LESSONS FROM A NATURAL VARIANT. *The Journal of Biological Chemistry,* 286**,** 15043-15049.

DOYE, J. P. K., LOUIS, A. A. & VENDRUSCOLO, M. 2004. Inhibition of protein crystallization by evolutionary negative design. *Physical Biology,* 1**,** P9-P13.

DUTTA, R. & INOUYE, M. 2000. GHKL, an emergent ATPase/kinase superfamily. *Trends in Biochemical Sciences,* 25**,** 24-28.

ELLIS, L., MORGAN, D. O., KOSHLAND, D. E., CLAUSER, E., MOE, G. R., BOLLAG, G., ROTH, R. A. & RUTTER, W. J. 1986. Linking Functional Domains of the Human Insulin-Receptor with the Bacterial Aspartate Receptor. *Proceedings of the National Academy of Sciences of the United States of America,* 83**,** 8137-8141.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallographica Section D,* 66**,** 486-501.

ERICSSON, U., HALLBERG, B., DETITTA, G., DEKKER, N. & NORDLUND, P. 2006. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry,* 357**,** 289-298.

ESSEN, L. O., MAILLIET, J. & HUGHES, J. 2008. The structure of a complete phytochrome sensory module in the Pr ground state. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 14709-14714.

ESTEBAN, B., CARRASCAL, M., ABIAN, J. & LAMPARTER, T. 2005. Light-induced conformational changes of cyanobacterial phytochrome Cph1 probed by limited proteolysis and autophosphorylation. *Biochemistry,* 44**,** 450-461.

EVANS, K., GROSSMANN, J. G., FORDHAM-SKELTON, A. P. & PAPIZ, M. Z. 2006. Small-angle X-ray scattering reveals the solution structure of a bacteriophytochrome in the catalytically active Pr state. *Journal of Molecular Biology,* 364**,** 655-666.

EVANS, P. 2006. Scaling and assessment of data quality. *Acta Crystallographica Section D,* 62**,** 72-82.

FENG, X. H., BAUMGARTNER, J. W. & HAZELBAUER, G. L. 1997. High- and low-abundance chemoreceptors in Escherichia coli: Differential activities associated with closely related cytoplasmic domains. *Journal of Bacteriology,* 179**,** 6714-6720.

FILONOV, G. S., PIATKEVICH, K. D., TING, L. M., ZHANG, J., KIM, K. & VERKHUSHA, V. V. 2011. Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nature Biotechnology,* 29**,** 757-761.

FRENCH, S. & WILSON, K. 1978. On the treatment of negative intensity observations. *Acta Crystallographica Section A,* 34**,** 517-525.

GIRAUD, E., ZAPPA, S., VUILLET, L., ADRIANO, J. M., HANNIBAL, L., FARDOUX, J., BERTHOMIEU, C., BOUYER, P., PIGNOL, D. & VERMEGLIO, A. 2005. A new type of bacteriophytochrome acts in tandem with a classical bacteriophytochrome to control the antennae synthesis in Rhodopseudomonas palustris. *J Biol Chem,* 280**,** 32389-32397.

GLIKO, O., NEUMAIER, N., FISCHER, M., HAASE, I., BACHER, A., WEINKAUF, S. & VEKILOV, P. G. 2005. Dense liquid droplets as a step source for the crystallization of lumazine synthase. *Journal of Crystal Growth,* 275**,** e1409-e1416.

GOLDSCHMIDT, L., COOPER, D. R., DEREWENDA, Z. S. & EISENBERG, D. 2007. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Science,* 16**,** 1569-1576.

GONZALEZ, C. V., IBARRA, S. E., PICCOLI, P. N., BOTTO, J. F. & BOCCALANDRO, H. E. 2012. Phytochrome B increases drought tolerance by enhancing ABA sensitivity in Arabidopsis thaliana. *Plant, cell & environment*.

GOODNER, B., HINKLE, G., GATTUNG, S., MILLER, N., BLANCHARD, M., QUROLLO, B., GOLDMAN, B. S., CAO, Y., ASKENAZI, M., HALLING, C., MULLIN, L., HOUMIEL, K., GORDON, J., VAUDIN, M., IARTCHOUK, O., EPP, A., LIU, F., WOLLAM, C., ALLINGER, M., DOUGHTY, D., SCOTT, C., LAPPAS, C., MARKELZ, B., FLANAGAN, C., CROWELL, C., GURSON, J., LOMO, C., SEAR, C., STRUB, G., CIELO, C. & SLATER, S. 2001. Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58. *Science,* 294**,** 2323-2328.

GUARNIERI, M. T., ZHANG, L., SHEN, J. & ZHAO, R. 2008. The Hsp90 inhibitor radicicol interacts with the ATP-binding pocket of bacterial sensor kinase PhoQ. *Journal of Molecular Biology,* 379**,** 82-93.

HAHN, J., STRAUSS, H. M., LANDGRAF, F. T., GIMENEZ, H. F., LOCHNIT, G., SCHMIEDER, P. & HUGHES, J. 2006. Probing protein-chromophore interactions in Cph1 phytochrome by mutagenesis. *FEBS Journal,* 273**,** 1415-1429.

HAHN, J., STRAUSS, H. M. & SCHMIEDER, P. 2008. Heteronuclear NMR investigation on the structure and dynamics of the chromophore binding pocket of the cyanobacterial phytochrome Cph1. *Journal of the American Chemical Society,* 130**,** 11170-11178.

HAMMOND, C. 2009. *The Basics of Crystallography and Diffraction,* Oxford university press.

HERDMAN, M., COURSIN, T., RIPPKA, R., HOUMARD, J. & TANDEAU DE MARSAC, N. 2000. A new appraisal of the prokaryotic origin of eukaryotic phytochromes. *Journal of molecular evolution,* 51**,** 205-213.

HESS, J. F., BOURRET, R. B. & SIMON, M. I. 1988. Histidine phosphorylation and phosphoryl group transfer in bacterial chemotaxis. *Nature,* 336**,** 139-143.

HOCH, J. A. & SILHAVY, T. J. 1995. *Two-Component Signal Transduction,* Washington, D.C., American Society for Microbiology.

HORITA, S., YAMANAKA, Y., YAMAMURA, A., OKADA, A., NAKAYAMA, J., NAGATA, K. & TANOKURA, M. 2011. Crystallization and preliminary X-ray analysis of a putative sensor histidine kinase domain: the C-terminal domain of HksP4 from Aquifex aeolicus VF5. *Acta Crystallographica Section F,* 67**,** 803-807.

HUANG, Y. & LIU, Z. 2009. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. *Journal of Molecular Biology,* 393**,** 1143-1159.

HUGHES, J. 2010. Phytochrome three-dimensional structures and functions. *Biochemical Society Transactions,* 38**,** 710-716.

HULKO, M., BERNDT, F., GRUBER, M., LINDER, J. U., TRUFFAULT, V., SCHULTZ, A., MARTIN, J., SCHULTZ, J. E., LUPAS, A. N. & COLES, M. 2006. The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell,* 126**,** 929-940.

IM, Y. J., RHO, S.-H., PARK, C.-M., YANG, S.-S., KANG, J.-G., LEE, J. Y., SONG, P.-S. & EOM, S. H. 2009. Crystal structure of a cyanobacterial phytochrome response regulator. *Protein Science,* 11**,** 614-624.

INOMATA, K., HAMMAM, M. A., KINOSHITA, H., MURATA, Y., KHAWN, H., NOACK, S., MICHAEL, N. & LAMPARTER, T. 2005. Sterically locked synthetic bilin derivatives and phytochrome Agp1 from Agrobacterium tumefaciens form photoinsensitive Pr- and Pfr-like adducts. *The Journal of Biological Chemistry,* 280**,** 24491-24497.

INOMATA, K., KHAWN, H., CHEN, L.-Y., KINOSHITA, H., ZIENICKE, B., MOLINA, I. & LAMPARTER, T. 2009. Assembly of Agrobacterium Phytochromes Agp1 and Agp2 with Doubly Locked Bilin Chromophores†. *Biochemistry,* 48**,** 2817-2827.

INOMATA, K., NOACK, S., HAMMAM, M. A., KHAWN, H., KINOSHITA, H., MURATA, Y., MICHAEL, N., SCHEERER, P., KRAUSS, N. & LAMPARTER, T. 2006. Assembly of synthetic locked chromophores with agrobacterium phytochromes Agp1 and Agp2. *J Biol Chem,* 281**,** 28162-28173.

JIANG, Z., SWEM, L. R., RUSHING, B. G., DEVANATHAN, S., TOLLIN, G. & BAUER, C. E. 1999. Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes. *Science,* 285**,** 406-409.

JOOSTEN, R. P., TE BEEK, T. A. H., KRIEGER, E., HEKKELMAN, M. L., HOOFT, R. W. W., SCHNEIDER, R., SANDER, C. & VRIEND, G. 2011. A series of PDB related databases for everyday needs. *Nucleic Acids Research,* 39**,** D411-D419.

KABSCH, W. 2010. XDS. *Acta Crystallographica Section D,* 66**,** 125-132.

KABSCH, W. & SANDER, C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* 22**,** 2577-2637.

KANTARDJIEFF, K. A. & RUPP, B. 2003. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Science,* 12**,** 1865-1871.

KARNIOL, B. & VIERSTRA, R. D. 2003. The pair of bacteriophytochromes from Agrobacterium tumefaciens are histidine kinases with opposing photobiological properties. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 2807-2812.

KARNIOL, B. & VIERSTRA, R. D. 2004. The HWE Histidine Kinases, a New Family of Bacterial Two-Component Sensor Kinases with Potentially Diverse Roles in Environmental Signaling. *Journal of Bacteriology,* 186**,** 445-453.

KARNIOL, B., WAGNER, JEREMIAH R., WALKER, JOSEPH M. & VIERSTRA, RICHARD D. 2005. Phylogenetic analysis of the phytochrome superfamily reveals distinct microbial subfamilies of photoreceptors. *Biochemical Journal,* 392**,** 103-116.

KATO, M., MIZUNO, T., SHIMIZU, T. & HAKOSHIMA, T. 1997. Insights into multistep phosphorelay from the crystal structure of the C-terminal HPt domain of ArcB. *Cell,* 88**,** 717-723.

KEENER, J. & KUSTU, S. 1988. Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NTRB and NTRC of enteric bacteria: roles of the conserved amino-terminal domain of NTRC. *Proceedings of the National Academy of Sciences of the United States of America,* 85**,** 4976-4980.

KEEP, N. H., BARNES, M., BARSUKOV, I., BADII, R., LIAN, L. Y., SEGAL, A. W., MOODY, P. C. & ROBERTS, G. C. 1997. A modulator of rho family G proteins, rhoGDI, binds these G proteins via an immunoglobulin-like domain and a flexible N-terminal arm. *Structure,* 5**,** 623-633.

KENNEY, L. J. 2010. How important is the phosphatase activity of sensor kinases? *Current Opinion in Microbiology,* 13**,** 168-176.

KRIKOS, A., CONLEY, M. P., BOYD, A., BERG, H. C. & SIMON, M. I. 1985. Chimeric Chemosensory Transducers of Escherichia-Coli. *Proceedings of the National Academy of Sciences of the United States of America,* 82**,** 1326-1330.

LAMPARTER, T. 2004. Evolution of cyanobacterial and plant phytochromes. *FEBS Letters,* 573**,** 1-5.

LAMPARTER, T., ESTEBAN, B. & HUGHES, J. 2001. Phytochrome Cph1 from the cyanobacterium Synechocystis PCC6803. Purification, assembly, and quaternary structure. *European Journal of Biochemistry,* 268**,** 4720-4730.

LAMPARTER, T., MICHAEL, N., CASPANI, O., MIYATA, T., SHIRAI, K. & INOMATA, K. 2003. Biliverdin binds covalently to agrobacterium phytochrome Agp1 via its ring A vinyl side chain. *The Journal of Biological Chemistry,* 278**,** 33786-33792.

LAMPARTER, T., MICHAEL, N., MITTMANN, F. & ESTEBAN, B. 2002. Phytochrome from Agrobacterium tumefaciens has unusual spectral properties and reveals an N-terminal chromophore attachment site. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 11628-11633.

LAMPARTER, T., MITTMANN, F., GARTNER, W., BORNER, T., HARTMANN, E. & HUGHES, J. 1997. Characterization of recombinant phytochrome from the cyanobacterium Synechocystis. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 11792-11797.

LARIMER, F. W., CHAIN, P., HAUSER, L., LAMERDIN, J., MALFATTI, S., DO, L., LAND, M. L., PELLETIER, D. A., BEATTY, J. T., LANG, A. S., TABITA, F. R., GIBSON, J. L., HANSON, T. E., BOBST, C., TORRES, J. L. T. Y., PERES, C., HARRISON, F. H., GIBSON, J. & HARWOOD, C. S. 2004. Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris. *Nature Biotechnology,* 22**,** 55-61.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and clustal X version 2.0. *Bioinformatics,* 23**,** 2947-2948.

LAUB, M. T. & GOULIAN, M. 2007. Specificity in two-component signal transduction pathways. *Annual Review of Genetics,* 41**,** 121-145.

LAWRENCE, M. C., MCKERN, N. M. & WARD, C. W. 2007. Insulin receptor structure and its implications for the IGF-1 receptor. *Current Opinion in Structural Biology,* 17**,** 699-705.

LAWSON, D. M., ARTYMIUK, P. J., YEWDALL, S. J., SMITH, J. M. A., LIVINGSTONE, J. C., TREFFRY, A., LUZZAGO, A., LEVI, S., AROSIO, P., CESARENI, G., THOMAS, C. D., SHAW, W. V. & HARRISON, P. M. 1991. Solving the Structure of Human H-Ferritin by Genetically Engineering Intermolecular Crystal Contacts. *Nature,* 349**,** 541-544.

LECOQ, J. & SCHNITZER, M. J. 2011. An infrared fluorescent protein for deeper imaging. *Nature Biotechnology,* 29**,** 715-716.

LEE, H. S., CHOI, J. & YOON, S. 2007. QHELIX: A computational tool for the improved measurement of inter-helical angles in proteins. *Protein Journal,* 26**,** 556-561.

LESLIE, A. G. W. & POWELL, H. R. 2007. Processing diffraction data with mosflm. *In:* READ, R. J. & SUSSMAN, J. L. (eds.) *Evolving Methods for Macromolecular Crystallography.* Springer Netherlands.

LETUNIC, I., DOERKS, T. & BORK, P. 2012. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research,* 40**,** D302-D305.

LEVSKAYA, A., CHEVALIER, A. A., TABOR, J. J., SIMPSON, Z. B., LAVERY, L. A., LEVY, M., DAVIDSON, E. A., SCOURAS, A., ELLINGTON, A. D., MARCOTTE, E. M. & VOIGT, C. A. 2005. Engineering Escherichia coli to see light - These smart bacteria 'photograph' a light pattern as a high-definition chemical image. *Nature,* 438**,** 441-442.

LEVSKAYA, A., WEINER, O. D., LIM, W. A. & VOIGT, C. A. 2009. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature,* 461**,** 997-1001.

LI, H., ZHANG, J. R., VIERSTRA, R. D. & LI, H. L. 2010. Quaternary organization of a phytochrome dimer as revealed by cryoelectron microscopy. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 10872-10877.

LISTWAN, P., TERWILLIGER, T. & WALDO, G. 2009. Automated, high-throughput platform for protein solubility screening using a split-GFP system. *Journal of structural and functional genomics,* 10**,** 47-55.

LONGENECKER, K. L., GARRARD, S. M., SHEFFIELD, P. J. & DEREWENDA, Z. S. 2001. Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Crystallographica Section D,* 57**,** 679-688.

LOVELL, S. C., DAVIS, I. W., ADRENDALL, W. B., DE BAKKER, P. I. W., WORD, J. M., PRISANT, M. G., RICHARDSON, J. S. & RICHARDSON, D. C. 2003. Structure validation by C alpha geometry: phi,psi and C beta deviation. *Proteins-Structure Function and Genetics,* 50**,** 437-450.

MAILLIET, J., PSAKIS, G., FEILKE, K., SINESHCHEKOV, V., ESSEN, L. O. & HUGHES, J. 2011. Spectroscopy and a high-resolution crystal structure of Tyr263 mutants of cyanobacterial phytochrome Cph1. *Journal of Molecular Biology,* 413**,** 115-127.

MARINA, A., MOTT, C., AUYZENBERG, A., HENDRICKSON, W. A. & WALDBURGER, C. D. 2001. Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. Insight into the reaction mechanism. *The Journal of Biological Chemistry,* 276**,** 41182-411890.

MARINA, A., WALDBURGER, C. D. & HENDRICKSON, W. A. 2005. Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *The EMBO Journal,* 24**,** 4247-4259.

MATEJA, A., DEVEDJIEV, Y., KROWARSCH, D., LONGENECKER, K., DAUTER, Z., OTLEWSKI, J. & DEREWENDA, Z. S. 2002. The impact of Glu-->Ala and Glu-->Asp mutations on the crystallization properties of RhoGDI: the structure of RhoGDI at 1.3 A resolution. *Acta Crystallographica Section D,* 58**,** 1983-1991.

MATSUSHITA, M. & JANDA, K. D. 2002. Histidine kinases as targets for new antimicrobial agents. *Bioorganic & medicinal chemistry,* 10**,** 855-867.

MATTHEWS, E. E., ZOONENS, M. & ENGELMAN, D. M. 2006. Dynamic helix interactions in transmembrane signaling. *Cell,* 127**,** 447-450.

MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *Journal of Applied Crystallography,* 40**,** 658-674.

MEHER, A. K., BLABER, S. I., LEE, J., HONJO, E., KUROKI, R. & BLABER, M. 2009. Engineering an improved crystal contact across a solvent-mediated interface of human fibroblast growth factor 1. *Acta Crystallographica Section F,* 65**,** 1136-1140.

MIZUTANI, H., SARABOJI, K., MALATHY SONY, S. M., PONNUSWAMY, M. N., KUMAREVEL, T., KRISHNA SWAMY, B. S., SIMANSHU, D. K., MURTHY, M. R. N. & KUNISHIMA, N. 2008. Systematic study on crystal-contact engineering of diphthine synthase: influence of mutations at crystal-packing regions on X-ray diffraction quality. *Acta Crystallographica Section D,* 64**,** 1020-1033.

MOE, G. R., BOLLAG, G. E. & KOSHLAND, D. E. 1989. Transmembrane Signaling by a Chimera of the Escherichia-Coli Aspartate Receptor and the Human Insulin-Receptor. *Proceedings of the National Academy of Sciences of the United States of America,* 86**,** 5683-5687.

MÖGLICH, A., AYERS, R. A. & MOFFAT, K. 2009. Design and Signaling Mechanism of Light-Regulated Histidine Kinases. *Journal of Molecular Biology,* 385**,** 1433-1444.

MOGLICH, A. & MOFFAT, K. 2010. Engineered photoreceptors as novel optogenetic tools. *Photochemical & Photobiological Sciences,* 9**,** 1286-1300.

MONTGOMERY, B. L. & LAGARIAS, J. C. 2002. Phytochrome ancestry: sensors of bilins and light. *Trends in Plant Science,* 7**,** 357-366.

MOUREY, L., DA RE, S., PEDELACQ, J. D., TOLSTYKH, T., FAURIE, C., GUILLET, V., STOCK, J. B. & SAMAMA, J. P. 2001. Crystal structure of the CheA histidine phosphotransfer domain that mediates response regulator phosphorylation in bacterial chemotaxis. *The Journal of Biological Chemistry,* 276**,** 31074-31082.

MOUSSATCHE, P. & KLEE, H. J. 2004. Autophosphorylation activity of the Arabidopsis ethylene receptor multigene family. *The Journal of Biological Chemistry,* 279**,** 48734-48741.

MURSHUDOV, G. N., SKUBAK, P., LEBEDEV, A. A., PANNU, N. S., STEINER, R. A., NICHOLLS, R. A., WINN, M. D., LONG, F. & VAGIN, A. A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D,* 67**,** 355-367.

NAGATANI, A. 2010. Phytochrome: structural basis for its functions. *Current Opinion in Plant Biology,* 13**,** 565-570.

NI, M., TEPPERMAN, J. M. & QUAIL, P. H. 1999. Binding of phytochrome B to its nuclear signalling partner PIF3 is reversibly induced by light. *Nature,* 400**,** 781-784.

NINFA, A. J., NINFA, E. G., LUPAS, A. N., STOCK, A., MAGASANIK, B. & STOCK, J. 1988. Crosstalk between Bacterial Chemotaxis Signal Transduction Proteins and Regulators of Transcription of the Ntr Regulon - Evidence That Nitrogen Assimilation and Chemotaxis Are Controlled by a Common Phosphotransfer Mechanism. *Proceedings of the National Academy of Sciences of the United States of America,* 85**,** 5492-5496.

NIXON, B. T., RONSON, C. W. & AUSUBEL, F. M. 1986. 2-Component Regulatory Systems Responsive to Environmental Stimuli Share Strongly Conserved Domains with the Nitrogen Assimilation Regulatory Genes Ntrb and Ntrc. *Proceedings of the National Academy of Sciences of the United States of America,* 83**,** 7850-7854.

NJIMONA, I. & LAMPARTER, T. 2011. Temperature effects on Agrobacterium phytochrome Agp1. *PloS one,* 6**,** e25977.

NOACK, S., MICHAEL, N., ROSEN, R. & LAMPARTER, T. 2007. Protein conformational changes of Agrobacterium phytochrome Agp1 during chromophore assembly and photoconversion. *Biochemistry,* 46**,** 4164-4176.

OBERPICHLER, I., MOLINA, I., NEUBAUER, O. & LAMPARTER, T. 2006. Phytochromes from Agrobacterium tumefaciens: difference spectroscopy with extracts of wild type and knockout mutants. *FEBS Letters,* 580**,** 437-442.

OBERPICHLER, I., PIERIK, A. J., WESSLOWSKI, J., POKORNY, R., ROSEN, R., VUGMAN, M., ZHANG, F., NEUBAUER, O., RON, E. Z., BATSCHAUER, A. & LAMPARTER, T. 2011. A photolyase-like protein from Agrobacterium tumefaciens with an iron-sulfur cluster. *PloS one,* 6**,** e26775.

OBERPICHLER, I., ROSEN, R., RASOULY, A., VUGMAN, M., RON, E. Z. & LAMPARTER, T. 2008. Light affects motility and infectivity ofAgrobacterium tumefaciens. *Environmental Microbiology,* 10**,** 2020-2029.

PADILLA, J. E. & YEATES, T. O. 2003. A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallographica Section D,* 59**,** 1124-1130.

PARASHAR, V., MIROUZE, N., DUBNAU, D. A. & NEIDITCH, M. B. 2011. Structural Basis of Response Regulator Dephosphorylation by Rap Phosphatases. *PLoS Biol,* 9**,** e1000589.

PARK, S. & CRANE, B. R. 2011. Crystallization and preliminary X-ray crystallographic analysis of CheW from Thermotoga maritima: a coupling protein of CheA and the chemotaxis receptor. *Acta Crystallographica Section F,* 67**,** 504-506.

PARK, S. Y., HAM, S. W., KIM, K. Y. & CRANE, B. R. 2011. Crystallization and preliminary X-ray crystallographic analysis of Escherichia coli CheA P3 dimerization domain. *Acta Crystallographica Section F,* 67**,** 662-664.

PARKINSON, J. S. & KOFOID, E. C. 1992. Communication Modules in Bacterial Signaling Proteins. *Annual Review of Genetics,* 26**,** 71-112.

PERRAUD, A. L., KIMMEL, B., WEISS, V. & GROSS, R. 1998. Specificity of the BvgAS and EvgAS phosphorelay is mediated by the C-terminal HPt domains of the sensor proteins. *Molecular microbiology,* 27**,** 875-87.

PETSEV, D. N., THOMAS, B. R., YAU, S. T., TSEKOVA, D., NANEV, C., WILLIAM WILSON, W. & VEKILOV, P. G. 2001. Temperature-independent solubility and interactions between apoferritin monomers and dimers in solution. *Journal of Crystal Growth,* 232**,** 21-29.

PRICE II, W. N., CHEN, Y., HANDELMAN, S. K., NEELY, H., MANOR, P., KARLIN, R., NAIR, R., LIU, J., BARAN, M., EVERETT, J., TONG, S. N., FOROUHAR, F., SWAMINATHAN, S. S., ACTON, T., XIAO, R., LUFT, J. R., LAURICELLA, A., DETITTA, G. T., ROST, B., MONTELIONE, G. T. & HUNT, J. F. 2008. Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nature Biotechnology,* 27**,** 51-57.

PSAKIS, G., MAILLIET, J., LANG, C., TEUFEL, L., ESSEN, L. O. & HUGHES, J. 2011. Signaling kinetics of cyanobacterial phytochrome Cph1, a light regulated histidine kinase. *Biochemistry,* 50**,** 6178-6188.

PUTHIYAVEETIL, S., KAVANAGH, T. A., CAIN, P., SULLIVAN, J. A., NEWELL, C. A., GRAY, J. C., ROBINSON, C., VAN DER GIEZEN, M., ROGERS, M. B. & ALLEN, J. F. 2008. The ancestral symbiont sensor kinase CSK links photosynthesis with gene expression in chloroplasts. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 10061-10066.

QUEST, B. & GÄRTNER, W. 2004. Chromophore selectivity in bacterial phytochromes. *European Journal of Biochemistry,* 271**,** 1117-1126.

RAYMENT, I., RYPNIEWSKI, W. R., SCHMIDTBASE, K., SMITH, R., TOMCHICK, D. R., BENNING, M. M., WINKELMANN, D. A., WESENBERG, G. & HOLDEN, H. M. 1993. Three-Dimensional Structure of Myosin Subfragment-1 - a Molecular Motor. *Science,* 261**,** 50-58.

ROCKWELL, N. C. & LAGARIAS, J. C. 2010. A Brief History of Phytochromes. *Chemphyschem,* 11**,** 1172-1180.

ROCKWELL, N. C., MARTIN, S. S., FEOKTISTOVA, K. & LAGARIAS, J. C. 2011. Diverse two-cysteine photocycles in phytochromes and cyanobacteriochromes. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 11854-11859.

ROCKWELL, N. C., SHANG, L., MARTIN, S. S. & LAGARIAS, J. C. 2009. Distinct classes of red/far-red photochemistry within the phytochrome superfamily. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 6123-6127.

ROCKWELL, N. C., SU, Y.-S. & LAGARIAS, J. C. 2006. Phytochrome Structure and Signaling Mechanisms. *Annual Review of Plant Biology,* 57**,** 837-858.

ROHMER, T., LANG, C., HUGHES, J., ESSEN, L. O., GARTNER, W. & MATYSIK, J. 2008. Light-induced chromophore activity and signal transduction in phytochromes observed by 13C and 15N magic-angle spinning NMR. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 15229-15234.

ROTTWINKEL, G., OBERPICHLER, I. & LAMPARTER, T. 2010. Bathy phytochromes in rhizobial soil bacteria. *Journal of Bacteriology,* 192**,** 5124-5133.

RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M. A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics (Oxford, England),* 16**,** 944-945.

SAFF, E. & KUIJLAARS, A. 1997. Distributing many points on a sphere. *The Mathematical Intelligencer,* 19**,** 5-11.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. 1989. *Molecular cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press.

SAMMA, A. A., JOHNSON, C. K., SONG, S., ALVAREZ, S. & ZIMMER, M. 2010. On the origin of fluorescence in bacteriophytochrome infrared fluorescent proteins. *J Phys Chem B,* 114**,** 15362-15369.

SCHALL, C. A., ARNOLD, E. & WIENCEK, J. M. 1996. Enthalpy of crystallization of hen egg-white lysozyme. *Journal of Crystal Growth,* 165**,** 293-298.

SCHEER, H. & ZHAO, K. H. 2008. Biliprotein maturation: the chromophore attachment. *Molecular microbiology,* 68**,** 263-276.

SCHEERER, P., MICHAEL, N., PARK, J. H., NAGANO, S., CHOE, H.-W., INOMATA, K., BORUCKI, B., KRAUß, N. & LAMPARTER, T. 2010. Light-Induced Conformational Changes of the Chromophore and the Protein in Phytochromes: Bacterial Phytochromes as Model Systems. *Chemphyschem,* 11**,** 1090-1105.

SCHEERER, P., MICHAEL, N., PARK, J. H., NOACK, S., FÖRSTER, C., HAMMAM, M. A. S., INOMATA, K., CHOE, H.-W., LAMPARTER, T. & KRAUß, N. 2006. Crystallization and preliminary X-ray crystallographic analysis of the N-terminal photosensory module of phytochrome Agp1, a biliverdin-binding photoreceptor from Agrobacterium tumefaciens. *Journal of Structural Biology,* 153**,** 97-102.

SCHÜTTELKOPF, A. W. & VAN AALTEN, D. M. F. 2004. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D,* 60**,** 1355-1363.

SHIMIZU-SATO, S., HUQ, E., TEPPERMAN, J. M. & QUAIL, P. H. 2002. A light-switchable gene promoter system. *Nature Biotechnology,* 20**,** 1041-1044.

SHU, X., ROYANT, A., LIN, M. Z., AGUILERA, T. A., LEV-RAM, V., STEINBACH, P. A. & TSIEN, R. Y. 2009. Mammalian expression of infrared fluorescent proteins engineered from a bacterial phytochrome. *Science,* 324**,** 804-807.

SKERKER, J. M., PERCHUK, B. S., SIRYAPORN, A., LUBIN, E. A., ASHENBERG, O., GOULIAN, M. & LAUB, M. T. 2008. Rewiring the specificity of two-component signal transduction systems. *Cell,* 133**,** 1043-1054.

SLABINSKI, L., JAROSZEWSKI, L., RODRIGUES, A. P. C., RYCHLEWSKI, L., WILSON, I. A., LESLEY, S. A. & GODZIK, A. 2007. The challenge of protein structure determination - lessons from structural genomics. *Protein Science,* 16**,** 2472-2482.

SMYTH, D. R., MROZKIEWICZ, M. K., MCGRATH, W. J., LISTWAN, P. & KOBE, B. 2003. Crystal structures of fusion proteins with large-affinity tags. *Protein Science,* 12**,** 1313-1322.

SONG, C., PSAKIS, G., LANG, C., MAILLIET, J., GÄRTNER, W., HUGHES, J. & MATYSIK, J. 2011. Two ground state isoforms and a chromophore D-ring photoflip triggering extensive intramolecular changes in a canonical phytochrome. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 3842-3847.

SONG, H. K., LEE, J. Y., LEE, M. G., MOON, J., MIN, K., YANG, J. K. & SUH, S. W. 1999. Insights into eukaryotic multistep phosphorelay signal transduction revealed by the crystal structure of Ypd1p from Saccharomyces cerevisiae. *Journal of Molecular Biology,* 293**,** 753-761.

SONG, Y., PEISACH, D., PIOSZAK, A. A., XU, Z. H. & NINFA, A. J. 2004. Crystal structure of the c-terminal domain of the two-component system W transmitter protein nitrogen regulator II (NRII; NtrB), regulator of nitrogen assimilation in Escherichia coli. *Biochemistry,* 43**,** 6670-6678.

STEIN, N. 2008. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography,* 41**,** 641-643.

STOCK, A. M., ROBINSON, V. L. & GOUDREAU, P. N. 2000. Two-component signal transduction. *Annual Review of Biochemistry,* 69**,** 183-215.

SUN, Q., YODA, K., SUZUKI, M. & SUZUKI, H. 2003. Vascular tissue in the stem and roots of woody plants can conduct light. *Journal of Experimental Botany,* 54**,** 1627-1635.

SZURMANT, H. & HOCH, J. A. 2010. Interaction fidelity in two-component signaling. *Current Opinion in Microbiology,* 13**,** 190-197.

TAN, P. H., CHU, V., STRAY, J. E., HAMLIN, D. K., PETTIT, D., WILBUR, D. S., VESSELLA, R. L. & STAYTON, P. S. 1998. Engineering the isoelectric point of a renal cell carcinoma targeting antibody greatly enhances scFv solubility. *Immunotechnology : an international journal of immunological engineering,* 4**,** 107-114.

TANAKA, T., SAHA, S. K., TOMOMORI, C., ISHIMA, R., LIU, D., TONG, K. I., PARK, H., DUTTA, R., QIN, L., SWINDELLS, M. B., YAMAZAKI, T., ONO, A. M., KAINOSHO, M., INOUYE, M. & IKURA, M. 1998. NMR structure of the histidine kinase domain of the E. coli osmosensor EnvZ. *Nature,* 396**,** 88-92.

TASLER, R., MOISES, T. & FRANKENBERG-DINKEL, N. 2005. Biochemical and spectroscopic characterization of the bacterial phytochrome of Pseudomonas aeruginosa. *FEBS Journal,* 272**,** 1927-1936.

TERWILLIGER, T. C., STUART, D. & YOKOYAMA, S. 2009. Lessons from Structural Genomics. *Annual Review of Biophysics,* 38**,** 371-383.

TOMOMORI, C., TANAKA, T., DUTTA, R., PARK, H., SAHA, S. K., ZHU, Y., ISHIMA, R., LIU, D., TONG, K. I., KUROKAWA, H., QIAN, H., INOUYE, M. & IKURA, M. 1999. Solution structure of the homodimeric core domain of Escherichia coli histidine kinase EnvZ. *Nature Structural Biology,* 6**,** 729-734.

TREVINO, S. R., SCHOLTZ, J. M. & PACE, C. N. 2007. Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *Journal of Molecular Biology,* 366**,** 449-460.

ULIJASZ, A. T., CORNILESCU, G., CORNILESCU, C. C., ZHANG, J., RIVERA, M., MARKLEY, J. L. & VIERSTRA, R. D. 2010. Structural basis for the photoconversion of a phytochrome to the activated Pfr form. *Nature,* 463**,** 250-254.

ULIJASZ, A. T. & VIERSTRA, R. D. 2011. Phytochrome structure and photochemistry: recent advances toward a complete molecular picture. *Current Opinion in Plant Biology,* 14**,** 498-506.

UTSUMI, R., BRISSETTE, R. E., RAMPERSAUD, A., FORST, S. A., OOSAWA, K. & INOUYE, M. 1989. Activation of Bacterial Porin Gene-Expression by a Chimeric Signal Transducer in Response to Aspartate. *Science,* 245**,** 1246-1249.

VARUGHESE, K. I., MADHUSUDAN, ZHOU, X. Z., WHITELEY, J. M. & HOCH, J. A. 1998. Formation of a novel four-helix bundle and molecular recognition sites by dimerization of a response regulator phosphotransferase. *Molecular Cell,* 2**,** 485-493.

VARUGHESE, K. I., TSIGELNY, I. & ZHAO, H. Y. 2006. The crystal structure of beryllofluoride Spo0F in complex with the phosphotransferase Spo0B represents a phosphotransfer pretransition state. *Journal of Bacteriology,* 188**,** 4970-4977.

VEKILOV, P. G. 2003. Solvent Entropy Effects in the Formation of Protein Solid Phases. *In:* CHARLES W. CARTER, JR. & ROBERT, M. S. (eds.) *Methods in Enzymology.* Academic Press.

VIERSTRA, R. D. & ZHANG, J. 2011. Phytochrome signaling: solving the Gordian knot with microbial relatives. *Trends in Plant Science,* 16**,** 417-426.

VU, A., HAMEL, D. J., ZHOU, H. & DAHLQUIST, F. W. 2011. The structure and dynamic properties of the complete histidine phosphotransfer domain of the chemotaxis specific histidine autokinase CheA from Thermotoga maritima. *Journal of biomolecular NMR,* 51**,** 49-55.

VUILLET, L., KOJADINOVIC, M., ZAPPA, S., JAUBERT, M., ADRIANO, J. M., FARDOUX, J., HANNIBAL, L., PIGNOL, D., VERMEGLIO, A. & GIRAUD, E. 2007. Evolution of a bacteriophytochrome from light to redox sensor. *The EMBO Journal,* 26**,** 3322-3331.

WAGNER, J. R., BRUNZELLE, J. S., FOREST, K. T. & VIERSTRA, R. D. 2005. A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature,* 438**,** 325-331.

WAGNER, J. R., ZHANG, J., BRUNZELLE, J. S., VIERSTRA, R. D. & FOREST, K. T. 2007. High resolution structure of Deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution. *J Biol Chem,* 282**,** 12298-12309.

WAGNER, J. R., ZHANG, J., VON STETTEN, D., GUNTHER, M., MURGIDA, D. H., MROGINSKI, M. A., WALKER, J. M., FOREST, K. T., HILDEBRANDT, P. & VIERSTRA, R. D. 2008. Mutational analysis of Deinococcus radiodurans bacteriophytochrome reveals key amino acids necessary for the photochromicity and proton exchange cycle of phytochromes. *The Journal of biological chemistry,* 283**,** 12212-12226.

WALLACE, A. C., LASKOWSKI, R. A. & THORNTON, J. M. 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering,* 8**,** 127-134.

WALTER, T. S., MEIER, C., ASSENBERG, R., AU, K. F., REN, J. S., VERMA, A., NETTLESHIP, J. E., OWENS, R. J., STUART, D. I. & GRIMES, J. M. 2006. Lysine methylation as a routine rescue strategy for protein crystallization. *Structure,* 14**,** 1617-1622.

WARD, S. M., DELGADO, A., GUNSALUS, R. P. & MANSON, M. D. 2002. A NarX-Tar chimera mediates repellent chemotaxis to nitrate and nitrite. *Molecular microbiology,* 44**,** 709-719.

WEERASURIYA, S., SCHNEIDER, B. M. & MANSON, M. D. 1998. Chimeric chemoreceptors in Escherichia coli: Signaling properties of Tar-Tap and Tap-Tar hybrids. *Journal of Bacteriology,* 180**,** 914-920.

WEISS, M. S. & HILGENFELD, R. 1997. On the use of the merging R factor as a quality indicator for X-ray data. *Journal of Applied Crystallography,* 30**,** 203-205.

WHITMORE, L. & WALLACE, B. A. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research,* 32**,** W668-W673.

WINN, M. D., BALLARD, C. C., COWTAN, K. D., DODSON, E. J., EMSLEY, P., EVANS, P. R., KEEGAN, R. M., KRISSINEL, E. B., LESLIE, A. G., MCCOY, A., MCNICHOLAS, S. J., MURSHUDOV, G. N., PANNU, N. S., POTTERTON, E. A., POWELL, H. R., READ, R. J., VAGIN, A. & WILSON, K. S. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D,* 67**,** 235-242.

WOOD, D. W., SETUBAL, J. C., KAUL, R., MONKS, D. E., KITAJIMA, J. P., OKURA, V. K., ZHOU, Y., CHEN, L., WOOD, G. E., ALMEIDA, N. F., WOO, L., CHEN, Y., PAULSEN, I. T., EISEN, J. A., KARP, P. D., BOVEE, D., CHAPMAN, P., CLENDENNING, J., DEATHERAGE, G., GILLET, W., GRANT, C., KUTYAVIN, T., LEVY, R., LI, M.-J., MCCLELLAND, E., PALMIERI, A., RAYMOND, C., ROUSE, G., SAENPHIMMACHAK, C., WU, Z., ROMERO, P., GORDON, D., ZHANG, S., YOO, H., TAO, Y., BIDDLE, P., JUNG, M., KRESPAN, W., PERRY, M., GORDON-KAMM, B., LIAO, L., KIM, S., HENDRICK, C., ZHAO, Z.-Y., DOLAN, M., CHUMLEY, F., TINGEY, S. V., TOMB, J.-F., GORDON, M. P., OLSON, M. V. & NESTER, E. W. 2001. The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58. *Science,* 294**,** 2317-2323.

WU, S.-H. & LAGARIAS, J. C. 2000. Defining the Bilin Lyase Domain: Lessons from the Extended Phytochrome Superfamily†. *Biochemistry,* 39**,** 13487-13495.

WU, S. J., LUO, J., O'NEIL, K. T., KANG, J., LACY, E. R., CANZIANI, G., BAKER, A., HUANG, M., TANG, Q. M., RAJU, T. S., JACOBS, S. A., TEPLYAKOV, A., GILLILAND, G. L. & FENG, Y. 2010. Structure-based engineering of a monoclonal antibody for improved solubility. *Protein engineering, design & selection : PEDS,* 23**,** 643-651.

XIE, W., DICKSON, C., KWIATKOWSKI, W. & CHOE, S. 2010. Structure of the cytoplasmic segment of histidine kinase receptor QseC, a key player in bacterial virulence. *Protein and peptide letters,* 17**,** 1383-1391.

XU, Q., PORTER, S. W. & WEST, A. H. 2003. The Yeast YPD1/SLN1 Complex. *Structure,* 11**,** 1569-1581.

XU, Q. P. & WEST, A. H. 1999. Conservation of structure and function among histidine-containing phosphotransfer (HPt) domains as revealed by the crystal structure of YPD1. *Journal of Molecular Biology,* 292**,** 1039-1050.

YAMADA, S., SUGIMOTO, H., KOBAYASHI, M., OHNO, A., NAKAMURA, H. & SHIRO, Y. 2009. Structure of PAS-Linked Histidine Kinase and the Response Regulator Complex. *Structure,* 17**,** 1333-1344.

YANG, R., NISHIYAMA, K., KAMIYA, A., UKAJI, Y., INOMATA, K. & LAMPARTER, T. 2012. Assembly of Synthetic Locked Phycocyanobilin Derivatives with Phytochrome in Vitro and in Vivo in Ceratodon purpureus and Arabidopsis. *The Plant cell,* 24**,** 1936-1951.

YANG, X., KUK, J. & MOFFAT, K. 2008. Crystal structure of Pseudomonas aeruginosa bacteriophytochrome: Photoconversion and signal transduction. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 14715-14720.

YANG, X., KUK, J. & MOFFAT, K. 2009. Conformational differences between the Pfr and Pr states in Pseudomonas aeruginosa bacteriophytochrome. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 15639-15644.

YANG, X., REN, Z., KUK, J. & MOFFAT, K. 2011. Temperature-scan cryocrystallography reveals reaction intermediates in bacteriophytochrome. *Nature,* 479**,** 428-432.

YANG, X., STOJKOVIC, E. A., KUK, J. & MOFFAT, K. 2007. Crystal structure of the chromophore binding domain of an unusual bacteriophytochrome, RpBphP3, reveals residues that modulate photoconversion. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 12571-12576.

YAU, S. T., PETSEV, D. N., THOMAS, B. R. & VEKILOV, P. G. 2000. Molecular-level thermodynamic and kinetic parameters for the self-assembly of apoferritin molecules into crystals. *Journal of Molecular Biology,* 303**,** 667-678.

ZAPF, J., SEN, U., MADHUSUDAN, HOCH, J. A. & VARUGHESE, K. I. 2000. A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. *Structure,* 8**,** 851-862.

ZHAO, X. D., COPELAND, D. M., SOARES, A. S. & WEST, A. H. 2008. Crystal structure of a complex between the phosphorelay protein YPD1 and the response regulator domain of SLN1 bound to a phosphoryl analog. *Journal of Molecular Biology,* 375**,** 1141-1151.

ZHU, Y. & INOUYE, M. 2003. Analysis of the role of the EnvZ linker region in signal transduction using a chimeric Tar/EnvZ receptor protein, Tez1. *The Journal of Biological Chemistry,* 278**,** 22812-22819.

ZIENICKE, B., CHEN, L. Y., KHAWN, H., HAMMAM, M. A. S., KINOSHITA, H., REICHERT, J., ULRICH, A. S., INOMATA, K. & LAMPARTER, T. 2011. Fluorescence of Phytochrome Adducts with Synthetic Locked Chromophores. *The Journal of Biological Chemistry,* 286**,** 1103-1113.